Экономико-математическая библиотека

ВВЕДЕНИЕ В НЕЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

Экономико-
математическая
библиотека

ВВЕДЕНИЕ В НЕЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

Под редакцией К.-Х. ЭЛЬСТЕРА

ПЕРЕВОД С НЕМЕЦКОГО

Под редакцией И.И. ЕРЕМИНА

МОСКВА "НАУКА"
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
1985

22.18
В24
УДК 519.6

EINFÜHRUNG IN DIE
NICHTLINEARE OPTIMIERUNG
PROF. DR. HABIL. K-H. ELSTER
DR. R.REINHARDT
Technische Hochschule Ilmenau
PROF. DR. M.SCHÄUBLE
DR. G.DONATH
Pädagogische Hochschule
"N. K.Krupskaja" Halle
LEIPZIG
BSB B.G.TEUBNER, 1977

Введение в нелинейное программирование/ Эльстер К.-Х., Рейнгардт Р., Шойбле М., Донат Г. / Пер. с нем. под ред. И.И.Еремина. — М.: Наука. Главная редакция физико-математической литературы, 1985. — 264 с.

Книга представляет собой перевод изданной в 1977 г. в ГДР монографии. Она содержит обширный материал по теории и методам математического программирования. Книгу отличает тщательность методических проработок, обилие примеров и интерпретаций, в частности геометрических.

Для специалистов по прикладной математике и экономической кибернетике, а также для инженеров.

© BSB B.G. Teubner Verlagsgesellschaft, Leipzig, 1977

© Издательство "Наука". Главная редакция физико-математической литературы, перевод на русский язык, предисловие редактора перевода, 1985

В $\frac{1702070000-084}{053\,(02)-85}$ 43-85

Предисловие редактора перевода 5
Предисловие 7

1. Основные положения 11

1.1. Постановка задачи 11
1.2. Примеры задач нелинейного программирования 16
1.3. Классические задачи оптимизации 17
1.3.1. Предварительное замечание 17
1.3.2. Теорема о разрешимости системы уравнений 18
1.3.3. Множители Лагранжа 19

2. Выпуклость 22

2.1. Выпуклые множества и выпуклые конусы 22
2.1.1. К понятию выпуклого множества и выпуклого конуса 22
2.1.2. Операции над выпуклыми множествами 26
2.1.3. Топологические свойства выпуклых множеств 40
2.1.4. Теоремы отделимости 48
2.1.5. Экстремальные точки и опорные гиперплоскости 54
2.2. Выпуклые функции 58
2.2.1. Определение и основные свойства выпуклых функций 58
2.2.2. О минимуме выпуклых функций 63
2.2.3. Непрерывность выпуклых функций 63
2.2.4. Дифференцируемость выпуклых функций 66
2.2.5. Субдифференциал выпуклых функций 71
2.2.6. Замкнутые функции 79
2.2.7. Выпуклые функции f: $\mathbf{R}^n \to \bar{\mathbf{R}}$ 83
2.3. Обобщения выпуклых функций 86
2.4. Системы выпуклых неравенств 94

3. Сопряженные функции 99

3.1. Сопряженные множества 99
3.2. Сопряженные функции 104
3.3. Примеры сопряженных функций 110
3.3.1. Аффинные функции 110
3.3.2. Положительно однородные функции 111
3.3.3. Кусочно-линейные функции 111
3.3.4. Функции, постоянные на полупространстве 112
3.3.5. Квадратичные формы 113
3.4. Сопряженные для дифференцируемых функций 114

4. Критерии оптимальности 119

4.1. Постановка задачи 119
4.2. Локальные критерии оптимальности 120

4.2.1. Локальные критерии оптимальности для задачи P 120
4.2.2. Локальная теория множителей Лагранжа 123
4.2.3. Условия регулярности для задачи P_1 133
4.2.4. Локальная теория множителей Лагранжа для задачи P_1^* 136
4.3. Глобальная теория множителей Лагранжа 139
4.4. Критерии седловой точки для функции Лагранжа 142
4.5. Итоговый обзор 143

5. Теория двойственности 145

5.1. Двойственные задачи математического программирования 145
5.2. Свойства функции чувствительности и двойственной целевой функции 146
5.3. Другие утверждения о функции чувствительности 149
5.4. Сильная теорема двойственности 150
5.5. Примеры 154
5.6. Обратные теоремы двойственности 158
5.7. Теоремы двойственности для частных задач 162
5.7.1. Задача линейного программирования 162
5.7.2. Задача квадратичного программирования 163
5.8. Теоремы двойственности для задач с дифференцируемыми выпуклыми функциями ограничений и цели 165
5.9. Обобщения теоремы двойственности Фенхеля 168
5.10. Теоремы о минимаксе и двойственность 175

6. Методы решения 185

6.1. Основные понятия 185
6.2. Методы решения задач без ограничений 189
6.3. Методы прямого поиска 197
6.4. Методы решения задач с ограничениями 213
6.5. Методы возможных направлений 218
6.6. Методы штрафов и барьеров 229
6.6.1. Предварительное замечание 229
6.6.2. Метод штрафов 230
6.6.3. Метод штрафов для задач выпуклого программирования 233
6.6.4. Метод барьеров 235
6.6.5. Метод барьеров для задач выпуклого программирования 238
6.6.6. Смешанный метод штрафов и барьеров 240
6.7. Метод секущих плоскостей 243
6.8. Методы решения и точечно-множественные отображения 249

Список литературы 256

Список цитируемых работ, опубликованных на русском языке 259

Дополнительный список работ по оптимизации, опубликованных на русском языке 259

Предметный указатель 262

Настоящая книга посвящена теории и методам решения задач нелинейного программирования, которые находят все более широкую сферу приложений. Как правило, задачи оптимизации, возникающие при проектировании технических устройств, совершенствовании технологических процессов, при управлении теми или иными системами, носят нелинейный характер. Возрастает роль таких задач и при моделировании экономических задач.

Среди монографий по нелинейному программированию, изданных на русском языке, данная книга выделяется своей учебной направленностью. Авторский коллектив, возглавляемый известным специалистом в области математического программирования профессором К.-Х. Эльстером, положил в основу книги курс лекций, который читался в течение ряда лет на факультете математики, вычислительной техники и экономической кибернетики в Высшей технической школе г. Ильменау (ГДР). Несмотря на относительно небольшой объем, книга содержит обширный материал по теории и методам решения нелинейных оптимизационных задач. Следует отметить методическое совершенство книги, оптимальное, на наш взгляд, сочетание простоты и строгости изложения, обилие несложных, но ярких примеров и геометрических интерпретаций.

Книга состоит из шести глав. Первые пять глав посвящены теории нелинейного программирования. Здесь приводится постановка задачи нелинейного программирования, обсуждаются задачи классической оптимизации (глава 1), подробно рассматриваются вопросы выпуклого анализа (глава 2), теории сопряженных функций (глава 3), выводятся условия оптимальности (глава 4) и детально исследуется проблема двойственности в нелинейном программировании (глава 5). При изучении данных вопросов от читателя требуется минимальная математическая подготовка в объеме программы технических вузов. В главе 6 рассматривается ряд идей, которые могут быть положены в основу построения методов решения задач нелинейного программирования. Приводится краткий обзор методов для задач без ограничений, излагаются методы одномерной минимизации, штрафных функций, возможных направлений, секущих плоскостей.

Построение материала позволяет лицам, интересующимся вопросами собственно нелинейного программирования, начинать чтение непосредственно с главы 4, привлекая по мере надобности необходимые математические сведения из предыдущих глав.

Стараясь сохранить оригинальный характер изложения, мы избегали при переводе и редактировании книги каких-либо существенных правок. Были лишь устранены замеченные неточности и из основного списка цитируемой литературы выделены в отдельный список те издания, которые имеются на русском языке. Мы сочли целесообразным также привести дополнительный список изданных в СССР монографий по методам оптимизации и смежным вопросам.

Перевод книги выполнен В.Д.Скариным при участии Г.П.Барановой.

Книга предназначается в первую очередь для студентов математических, экономических и технических специальностей, изучающих теорию и методы задач оптимизации. Она будет, безусловно, полезна также многочисленным специалистам-практикам в области планирования, управления и проектирования.

И.И. Еремин

Настоящая книга посвящена нелинейному программированию — относительно молодой, но актуальной математической дисциплине. Она предназначается прежде всего для студентов математических, технических и экономических специальностей, а также для инженеров, занимающихся практическими оптимизационными задачами. Книга носит вводный характер и обязана своим возникновением лекциям, которые читались авторами на протяжении ряда лет студентам, специализирующимся в области математических методов исследования операций в Высшей технической школе Ильменау.

Предметом нелинейного программирования является класс экстремальных задач с ограничениями в форме равенств и неравенств. Такие задачи вызывают интерес у математиков, экономистов и инженеров в связи, например, с вопросами моделирования экономических, технических и технологических процессов, выработки оптимальных вариантов перевозок, распределения продуктов, определения последовательности обработки деталей. При этом исследованию подлежат проблема существования решений, структура множества решений и эффективные алгоритмы отыскания решений.

Первые работы по теории математического программирования в сегодняшнем понимании этих слов были опубликованы в 1939 г. Л.В. Канторовичем. В 1951 г. американские математики Г. Кун и А. Таккер заложили основы теории нелинейного программирования, обобщив классический принцип Эйлера — Лагранжа решения гладких экстремальных задач с ограничениями-равенствами.

Необходимость исследования разнообразных задач планирования и управления в хозяйственной и военной областях стимулировала бурное развитие математического программирования, которое продолжается и поныне. В настоящее время нелинейное программирование находит применение в самых различных постановках в экономике, а также в естественных и технических науках.

Нелинейное программирование использует аппарат многих математических дисциплин, прежде всего функционального анализа, топологии и алгебры. В свою очередь нелинейное программирование неоднократно существенно обогащало сами эти дисциплины. Вообще, развитие математики во многом стимулировало использование экстремальных задач.

Простые оптимизационные задачи геометрической природы были предметом глубокого изучения еще в античной Греции. Развитие дифференциального и интегрального исчисления в XVII и XVIII веках происходило

в тесной взаимосвязи с разработкой экстремальных принципов классической физики (таких, например, как принцип Ферма в геометрической оптике, вариационные принципы Эйлера – Лагранжа в механике, – последние позднее были вновь открыты и развиты Гамильтоном) и с возникновением вариационного исчисления как относительно самостоятельного раздела классического анализа. Современное развитие нелинейного программирования характеризуется двумя противоборствующими тенденциями:

1) исследованием все более специальных классов задач (в соответствии с разнообразными потребностями приложений);

2) возникновением единой обширной теории экстремальных задач в общих пространствах.

Анализ давно известных частных результатов привел к созданию (в значительной степени благодаря усилиям советских ученых) общих концепций единого подхода к широкому классу различных экстремальных задач – от классического вариационного исчисления до задач оптимального управления. Центральное место в этой конструкции занимает общая теория выпуклых множеств и операторов. Дело в том, что многие возникающие на практике модели являются выпуклыми задачами нелинейного программирования или могут быть различными способами аппроксимированы подобными задачами.

Вследствие этого основная часть книги посвящена вопросам выпуклого программирования. Теория выпуклых множеств и функций образует ядро книги, на основе которого в последующих главах исследуются задачи выпуклого программирования (и некоторые более общие классы задач).

При этом в основу рассмотрения положены задачи с вещественными выпуклыми функциями (целевой и ограничений), которые определены на подмножествах евклидова пространства $\mathbf{R}^n$. Авторы сознательно отказались от более общего представления по следующим двум причинам:

1. Во-первых, выпуклая оптимизация в $\mathbf{R}^n$ представляет собой относительно завершенный раздел математического программирования, при изучении которого читатель может ознакомиться со спецификой всей данной дисциплины. Для понимания изложенного материала необходимы лишь немногие предварительные сведения из линейной алгебры и дифференциального исчисления.

2. Во-вторых, на основании приведенной теории могут быть проанализированы многие возникающие на практике задачи. Кроме того, численное решение задач оптимизации в общих пространствах с помощью ЭВМ, как правило, требует предварительной "дискретизации" задач, т.е. сведения их к нелинейной оптимизационной задаче в $\mathbf{R}^n$. И наконец, задачи оптимизации в общих пространствах рассматриваются в книге А. Гепферта "Математическое программирование в общих векторных пространствах", вышедшей в этой же серии.

Объем книги накладывает, естественно, ограничения на освещение некоторых существенных разделов теории. Читатели, заинтересовавшиеся специальными вопросами, будут отосланы к соответствующей литературе. Но и здесь ввиду обилия публикаций авторы ограничились небольшим числом важнейших монографий и оригинальных статей.

Книга разбита на 6 глав, из которых первые три носят подготовительный характер.

Глава 1 содержит наряду с постановкой задачи соглашения о применяемой терминологии, вспомогательные сведения из дифференциального исчисления функций многих переменных, а также классическую теорию множителей Лагранжа для задач на экстремум без ограничений.

Глава 2 посвящена изложению основ теории выпуклых множеств в $\mathbf{R}^n$, выпуклых функций, а также некоторых обобщений выпуклых функций. Объем излагаемого материала превосходит необходимый для понимания последующих рассуждений уровень, но при этом авторы не ставили цели подменить имеющиеся монографии по выпуклому анализу. Читателю, интересующемуся специально вопросами оптимизации, можно порекомендовать начинать чтение книги с главы 4, привлекая по мере необходимости вспомогательные сведения из глав 2 и 3.

Глава 3 содержит изложение теории сопряженных (по Фенхелю) функций; в ней также приведены многочисленные примеры вычисления сопряженных функций.

Собственно теория выпуклого программирования рассмотрена в главах 4 и 5.

Глава 4, названная "Критерии оптимальности", знакомит читателя с локальной теорией множителей Лагранжа, а также с глобальными условиями оптимальности. Условия такого рода позволяют глубоко проанализировать структуру множеств решений рассматриваемых задач оптимизации.

Критериями оптимальности специального типа являются также содержащиеся в главе 5 теоремы двойственности. Читателю предлагаются три подхода к теории двойственности: один из них тесно связан с понятием устойчивости задачи оптимизации, второй подход основывается на применении теории сопряженных функций, а третий использует связь между теоремами двойственности в выпуклом программировании и теоремами о минимаксе.

Наконец, глава 6 посвящена методам нелинейного программирования, точнее, теории некоторых прототипов методов, играющих важную роль в приложениях. Это методы прямого поиска, методы возможных направлений, методы штрафов и барьеров, а также методы секущих плоскостей. Приведен краткий обзор методов для задач без ограничений. В прикладном плане действенность нелинейного программирования зависит прежде всего от существования эффективных методов решения задач оптимизации и от наличия производительных ЭВМ, на которых реализуются эти методы. Отсюда вытекает, что зачастую вопрос о целесообразности применения того или иного метода следует решать не только в зависимости от заданного типа задач, но и с учетом имеющейся вычислительной техники. Это требует, помимо опыта работы с ЭВМ, более глубокого знакомства с существующей в большом числе оригинальной литературой.

Практическая реализация алгоритмов вышла бы за рамки нашей книги. Далее сознательно не излагается единая теория методов и лишь в заключение приводится набросок возможного подхода к построению такой теории на основе точечно-множественных отображений.

При работе над рукописью авторы пользовались поддержкой сотрудников исследовательских групп "Оптимизация" Высшей технической школы Ильменау и Высшей педагогической школы им. Н. К. Крупской в Галле.

Особой благодарности заслуживают г-жа Борхардт, д-р Хофман и Шаде за помощь при подготовке рукописи, а также проф. д-р А.Гепферт и д-р Р.Нэзе за ценные замечания и советы. За тщательность при перепечатке рукописи и изготовлении рисунков авторы благодарят соответственно г-жу Б.Шенефельд и г-жу В.Бернер.

Сотрудникам издательства "Teubner" в Лейпциге авторы признательны за содействие в осуществлении намеченного и за оформление книги, а также за понимание при возникавших затруднениях со сроками.

Авторы

1.1. Постановка задачи

За последние три десятилетия математическое программирование выделилось в специальный раздел математики. Этот раздел на сегодняшний день состоит в свою очередь из нескольких дисциплин: линейного программирования, нелинейного программирования, теории оптимального управления, дискретного, параметрического, стохастического программирования и т.д. Наблюдаемое при этом многообразие задач оптимизации связано, помимо всего прочего, с тем, что математическое программирование зародилось непосредственно из рассмотрения практических постановок и поэтому отражает многообразие таких задач. Тенденция развития за последнее десятилетие состоит в создании единой теории экстремальных задач и установлении связи между результатами этой теории и соответствующими разделами классической математики, такими, например, как вариационное исчисление (см. книгу Иоффе и Тихомирова [96]).

Цель последующего изложения состоит в том, чтобы ввести читателя в проблематику нелинейного программирования. При этом надо учитывать, что нелинейное программирование за отмеченный промежуток времени интенсивно развивалось и развивается и поныне. Это развитие характеризуется следующими моментами:

а) возрастающей абстракцией – переходом от задач в n-мерном евклидовом пространстве к задачам в общих пространствах, от частных вопросов к более общим проблемам (см., например, книгу Гепферта [35]);

б) многообразием методов решения – разработкой большого числа алгоритмов для задач нелинейного программирования, что обусловлено разнообразием и спецификой задач при различных математических предположениях;

в) специализацией – исследованием специальных нелинейных задач в квадратичном программировании, геометрическом программировании и т.д.;

г) многообразием приложений – построением нелинейных моделей, охватывающих все более сложные ситуации. При этом следует отметить, с одной стороны, преимущество таких моделей по сравнению с линейными в точности отражения сути явлений, а с другой – недостаток в виде возросших теоретических и вычислительных трудностей. Представление о прикладных возможностях нелинейного программирования можно составить из книги Пуна [72].

Мы ограничимся в дальнейшем рассмотрением задач нелинейного программирования в вещественном n-мерном евклидовом пространстве $\mathbf{R}^n$.

Вообще, любую задачу, в которой разыскивается максимум или минимум функции одной или нескольких переменных на некотором множестве, можно считать задачей математического программирования. При этом ограничения на переменные задаются обычно в форме уравнений или неравенств, но могут также представлять собой, например, требования целочисленности. Подобные задачи давно интересуют математиков, физиков, химиков, инженеров и экономистов. Определенные типы задач математического программирования могли быть решены с помощью классического дифференциального и вариационного исчисления. Следует обратить внимание на то, что при этом речь шла обычно о решении задачи в принципе и что возникавшие вычислительные трудности часто ограничивали и даже сводили на нет непосредственное применение теоретических результатов. Лишь благодаря современным ЭВМ эта классическая "теория оптимизации" получила новые области приложений.

За прошедшие десятилетия появились, однако, задачи (прежде всего в области экономики), которые уже не могли быть решены с помощью известных классических методов. Это привело к необходимости создания нового математического аппарата. Среди прочего возникли линейное и нелинейное программирование.

Прежде чем сформулировать общую задачу нелинейного программирования, мы рассмотрим несколько примеров.

При изготовлении определенного продукта естественным из экономических соображений является требование максимизировать прибыль или минимизировать издержки производства, при этом необходимые ресурсы (рабочая сила, оборудование, сырье, полуфабрикаты) не могут превзойти заданных величин. Уже здесь налицо существенное сужение постановки задачи, поскольку желательно, чтобы одновременно минимизировались издержки производства, время изготовления и максимизировалась степень механизации и автоматизации производственного процесса. Таким образом, в этом случае имеется несколько критериев оптимальности.

Такие задачи с несколькими целевыми функциями (задачи векторной оптимизации) в настоящей книге рассматриваться не будут.

Пример 1.1. Пусть необходимые для производства x единиц продукта A затраты выражаются следующим образом:

$$K(x) = ax^k, \quad x > 0, \quad a > 0, \quad k > 1.$$

Здесь общие издержки, начиная с некоторого x_0, растут быстрее, чем число произведенных единиц продукта A (прогрессия издержек). Очевидно, что это экономически нежелательная ситуация.

Если, напротив, издержки производства продукта A определены формулой

$$K(x) = ax^k, \quad x > 0, \quad a > 0, \quad 0 < k < 1,$$

то общие затраты растут медленнее, чем число произведенных единиц A (дегрессия издержек). Это – чаще всего встречающаяся ситуация (эффект крупносерийного производства).

Наконец, в редко встречающемся случае регрессии издержек справедливо соотношение

$$K(x) = a - bx^k, \quad x > 0, \quad a, b, k > 0.$$

Если в общем случае затраты на производство x_i единиц продукта A_i ($i = 1, ..., n$) на некотором предприятии заданы в виде

$$K(x_1, ..., x_n) = \sum_{i=1}^{n} a_i x_i^{k_i}, \quad x_i > 0, \quad i = 1, ..., n,$$

где a_i, $k_i > 0$, а продукты A_i должны быть произведены в количестве, не меньшем b_i ($i = 1, ..., n$), и должны удовлетворять условию на ассортимент $\sum_{i=1}^{n} d_i x_i \geqslant d$, где d_i, $d > 0$, то минимизация издержек на предприятии означает, что требуется найти минимум, вообще говоря, нелинейной функции K на множестве

$$G = \{(x_1, ..., x_n)^t: \sum_{i=1}^{n} d_i x_i \geqslant d, \ x_i \geqslant b_i, \ i = 1, ..., n\}.$$

Пример 1.2. В методе наименьших квадратов для сглаживания результатов наблюдений $(\tilde{x}_i, \tilde{y}_i)$, $i = 1, ..., m$, применяют целую рациональную функцию n-й степени вида $f(x) = \sum_{j=0}^{n} a_j x^j$. При этом требуется найти минимум функции

$$Q(a_0, a_1, ..., a_n) = \sum_{i=1}^{m} (\tilde{y}_i - \sum_{j=0}^{n} a_j \tilde{x}_i^j)^2, \quad a_0, a_1, ..., a_n \in \mathbf{R}.$$

В данном случае речь идет о задаче математического программирования с квадратичной целевой функцией.

Пример 1.3. Пусть заданы замкнутый шар

$$\bar{S}(0, \rho) = \{x \in \mathbf{R}^3: \ x = (x_1, x_2, x_3)^t, \ \|x\| \leqslant \rho\}$$

радиуса $\rho > 0$ и точка $\bar{x} \in \mathbf{R}^3$, для которой $\|\bar{x}\| > \rho$. Проблема отыскания точки $x^0 \in \bar{S}(0, \rho)$, минимально удаленной от $\bar{x}$, приводит к следующей задаче: найти минимум функции

$$f(x_1, x_2, x_3) = (x_1 - \bar{x}_1)^2 + (x_2 - \bar{x}_2)^2 + (x_3 - \bar{x}_3)^2, \quad (x_1, x_2, x_3)^t \in \mathbf{R}^3,$$

при ограничении

$$x_1^2 + x_2^2 + x_3^2 \leqslant \rho^2.$$

Это пример задачи выпуклого программирования.

Пример 1.4. Из множества

$$G = \{(x, y)^t: \ |x| \leqslant 1, \quad |y| \leqslant 1\} \subseteq \mathbf{R}^2$$

требуется выбрать n точек $x^i = (x_i, y_i)^t$, $i = 1, ..., n$, таким образом, чтобы наименьшее расстояние между двумя такими точками было максимально. Эта задача сводится к задаче математического программирования: найти максимум $f(x^1, ..., x^n) = \min_{\substack{i, j = 1, ..., n \\ i \neq j}} \|x^i - x^j\|$ на множестве $G^n = G \times G \times ... \times G$.

Рассмотренные в примерах 1.1 – 1.4 задачи являются частными случаями следующей общей постановки. Под *общей задачей математического программирования* мы понимаем задачу

P_A: заданы множество $G \subseteq \mathbf{R}^n$ и функция

f: $D(f) \to \mathbf{R}$, $D(f) \supseteq G$, $D(f) \neq \phi$;

требуется найти

$$\alpha = \inf_{x \in G} f(x), \tag{1.1}$$

$$G^0 = \{x \in G: f(x) = \alpha\}. \tag{1.2}$$

Множество G называют *допустимым множеством* (*допустимой областью*) для задачи P_A, каждый элемент $x \in G$ – *допустимой точкой* задачи.

Если $G \neq \phi$, то задача P_A называется *совместной*. В противном случае задача *несовместна*. Функция f называется *целевой функцией*. Если существует допустимая точка $x^0 \in G$, для которой $\alpha = f(x^0)$, то x^0 – *решение* (*оптимальное решение*) задачи P_A. Соответственно множество G^0 называется *множеством решений* задачи.

Целевая функция f, называемая часто также критерием качества, является математическим описанием цели, к которой стремятся в исследуемом процессе; она делает возможными количественные сравнения. Допустимое множество G выражает условия, при которых протекает процесс. В литературе, ориентированной на приложения, для целевой функции можно найти, в зависимости от постановки задачи, и другие названия, такие как стоимостная функция, функция выигрыша и др.

Краткости ради мы условимся о следующем способе записи сформулированной выше задачи:

$$P_A: \min\{f(x): x \in G\}, \text{ *)} \tag{1.3}$$

где $f: D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$, $D(f) \neq \phi$. Кроме того, считаем $D(f) \supseteq G$. Для x^0 – решения задачи P_A – имеет место

$$f(x) \geqslant f(x^0) \quad \forall x \in G,$$

а потому мы говорим иногда также о *глобальном решении* этой задачи.

Допустимая точка $\hat{x} \in G$ называется *точкой локального минимума* f на множестве G, если существует окрестность $U(\hat{x})$ точки $\hat{x}$ такая, что

$$f(x) \geqslant f(\hat{x}) \quad \forall x \in G \cap U(\hat{x}).$$

Если в этом соотношении для $x \neq \hat{x}$ равенство не может иметь места, то $\hat{x}$ называется *точкой собственного локального минимума* f на G:

$$f(x) > f(\hat{x}) \quad \forall x \in (G \cap U(\hat{x})) \setminus \{\hat{x}\}.$$

Итак, в сформулированной задаче математического программирования P_A цель состоит в том, чтобы отыскать глобальные решения задачи и найти тем самым глобальный минимум f на множестве G.

Из соотношения $\inf_{x \in G} f(x) = -\sup_{x \in G} \{-f(x)\}$ следует, что можно огра-

---

*) Запись $\min\{f(x): x \in G\}$ служит для обозначения задач, в которых требуется определить инфимум или минимум (если он существует) функции f на множестве G, а также множество решений. Таким образом, задачу $\min\{f(x): x \in G\}$ следует отличать от чисел $\inf_{x \in G} f(x)$ или $\min_{x \in G} f(x)$.

ничиться задачами вида (1.3) (т.е. *задачами минимизации*) и не рассматривать отдельно *задачи максимизации*.

Для дальнейших исследований необходимо конкретизировать природу множества G.

Часто встречаются задачи, в которых $G \subseteq \mathbf{R}^n$ задано следующим образом:

$$G = \{x \in \mathbf{R}^n:\ x \in X;\ g_i(x) \leqslant 0,\ i \in I;\ h_j(x) = 0,\ j \in I^*\}, \tag{1.4}$$

где $I = \{1, ..., m\}$; $I^* = \{1, ..., p\}$, $g_i: D(g_i) \to \mathbf{R}$, $h_j: D(h_j) \to \mathbf{R}$,

$D(g_i), D(h_j) \subseteq \mathbf{R}^n$, $i \in I$, $j \in I^*$, $X \subseteq (\bigcap\limits_{i \in I} D(g_i)) \cap (\bigcap\limits_{j \in I^*} D(h_j))$, $X \neq \varphi$.

Допускается также, что I и (или) I^* – пустые множества.

Вводя векторные функции $g = (g_1, ..., g_m)'$; $h = (h_1, ..., h_p)'$, вместо (1.4) получим

$$G = \{x \in \mathbf{R}^n:\ x \in X,\ g(x) \leqslant 0,\ h(x) = 0\}. \tag{1.5}$$

Так как

$$g_i(x) \leqslant 0 \iff -g_i(x) \geqslant 0,$$

то можно, очевидно, ограничиться знаком неравенства, выбранным в (1.4) и (1.5). Фигурирующие в (1.4) функции g_i, $i \in I$, и h_j, $j \in I^*$, и соответствующие вектор-функции g и h называются *функциями ограничений*, соотношения $g_i(x) \leqslant 0$, $h_j(x) = 0$ и $g(x) \leqslant 0$, $h(x) = 0$ – *ограничениями* задачи P_A.

Если допустимая область G задачи P_A состоит из точек $\mathbf{R}^n$ с целочисленными координатами, то говорят о *задаче целочисленного программирования*. Такие задачи в дальнейшем рассматриваться не будут. Изложение целочисленного программирования (ориентированное на приложения) можно найти в книге Корбута и Финкельштейна [48].

Если множество G в P_A имеет вид (1.4) или (1.5), то имеем *задачу нелинейного программирования*

$$P_A^*:\ \min\{f(x):\ x \in G\},$$

$$G = \{x \in \mathbf{R}^n:\ x \in X;\ g_i(x) \leqslant 0,\ i \in I;\ h_j(x) = 0,\ j \in I^*\}. \tag{1.6}$$

При этом предполагается, что по меньшей мере одна из функций f, g_i, $i \in I$, или h_j, $j \in I^*$, не является аффинной *).

Задача P_A^* называется *задачей линейного программирования* (и обозначается P_L), если $f: \mathbf{R}^n \to \mathbf{R}$ – аффинная функция; $X = \mathbf{R}^n$; $D(g_i) = D(h_j) = \mathbf{R}^n$ $\forall i \in I$, $\forall j \in I^*$; $g(x) = Ax - b$; $h(x) = Bx - \bar{b}$; A, B – матрицы порядков $m \times n$ и $p \times n$ соответственно; $b \in \mathbf{R}^m$, $\bar{b} \in \mathbf{R}^p$.

Следовательно, задача линейного программирования имеет вид

$$P_L:\ \min\{c'x + c_0:\ x \in G\},$$
$$G = \{x \in \mathbf{R}^n:\ Ax \leqslant b,\ Bx = \bar{b}\}. \tag{1.7}$$

---

*) Функция $f: \mathbf{R}^n \to \mathbf{R}$ называется *аффинной*, если $f(x) = a'x + b$, где $a \in \mathbf{R}^n$, $b \in \mathbf{R}$.

Часто в записи допустимой области в явной форме присутствует требование $x \in \mathbf{R}^n_+ = \{x \in \mathbf{R}^n\colon x_i \geqslant 0, i = 1, \ldots, n\}$. Ограничения $x_i \geqslant 0, i = 1, \ldots, n$, называются *условиями неотрицательности.*

В связи с исследованием задач математического программирования встают вопросы о

1) существовании решения,
2) единственности решения,
3) необходимых условиях существования решения,
4) достаточных условиях существования решения;

здесь решение мы понимаем в определенном выше смысле. Кроме того, эти вопросы имеют значение также и для точек $x \in G$, в которых f достигает на G локального минимума.

1. 2. Примеры задач нелинейного программирования

В приложениях вызывают интерес прежде всего *частные* задачи нелинейного программирования, которые получаются из P_A при надлежащей конкретизации множеств G и $D(f)$, а также функции f.

Наряду с задачами линейного программирования имеют важное значение следующие классы задач.

Задачи классической оптимизации P_{KL} характеризуются тем, что $G = \mathbf{R}^n$ (*задачи на экстремум без ограничений, задачи "свободной минимизации"*) или $I = \phi$, $I^* \neq \phi$ (*экстремальные задачи с ограничениями в форме равенств*).

Подробнее задачи классической оптимизации с дифференцируемыми функциями мы рассмотрим в § 1.3.

Задачи выпуклого программирования обладают свойством: X, $D(f)$, $D(g_i)$ $(i \in I)$, $D(h_j)$ $(j \in I^*)$ – выпуклые множества, f, g_i – выпуклые, а h_j – аффинные функции.

Для задач выпуклого программирования, как это станет видно из последующих глав, имеется относительно цельная теория.

Задачи квадратичного программирования:

$$f(x) = x'Qx + c'x + c_0,$$

$$D(f) = D(g_i) = D(h_j) = \mathbf{R}^n \quad \forall i \in I \quad \forall j \in I^*,$$

$$g(x) = Ax - b, \quad h(x) = Bx - d,$$

где Q – симметрическая матрица порядка $n \times n$, $c \in \mathbf{R}^n$, $c_0 \in \mathbf{R}$, A – матрица порядка $m \times n$; $b \in \mathbf{R}^m$; B – матрица порядка $p \times n$; $d \in \mathbf{R}^p$.

Если Q положительно определена или положительно полуопределена, то имеем *задачу выпуклого квадратичного программирования.*

Задачи гиперболического программирования:

$$f(x) = z(x)/n(x),$$

где

$$z\colon D(z) \to \mathbf{R}, \quad D(z) \subseteq \mathbf{R}^n,$$

$$n\colon D(n) \to \mathbf{R}, \quad D(n) \subseteq \mathbf{R}^n, \quad n(x) > 0 \quad \forall x \in G.$$

В литературе (см., например, [34]) исследовались частные случаи этой задачи.

Другой часто встречающийся в приложениях и относительно хорошо изученный класс задач составляют задачи сепарабельного программирования. Этот класс характеризуется тем, что входящие в задачу функции являются сепарабельными, т.е. функциями, которые могут быть представлены как сумма некоторых функций, зависящих лишь от одной переменной. Особое значение имеют сепарабельные задачи с аффинными функциями ограничений.

Задачи сепарабельного программирования:

$$f(x) = \sum_{i=1}^{n} f_i(x_i), \quad D(f) = D(f_1) \times \ldots \times D(f_n),$$

$$g_i(x) = \sum_{k=1}^{n} g_{ik}(x_k), \quad D(g_i) = D(g_{i1}) \times \ldots \times D(g_{in}), \qquad i \in I,$$

h_j $(j \in I^*)$ – аффинные функции.

Следующий важный класс задач математического программирования – это задачи так называемого геометрического программирования (см. [4, 67]), среди которых выделяются задачи позиномиального программирования:

$$f = f_0; \quad g_i = f_i - 1, \quad i \in I,$$

где

$$f_i(x) = \sum_{k \in J[i]} c_k x_1^{a_{k1}} x_2^{a_{k2}} \ldots x_n^{a_{kn}}, \quad i \in I \cup \{0\},$$

$$J[i] = \{m_{i-1} + 1, \ldots, m_i\}, \quad m_{-1} = 0, \quad m_n = p,$$

$$m_{i-1} < m_i \quad \forall i \in I \cup \{0\},$$

$$c_k > 0; \; a_{kj} \in \mathbf{R}, \quad k \in \{1, \ldots, p\}, \quad j \in \{1, \ldots, n\},$$

$$D(f) = D(g_i) = \{x \in \mathbf{R}^n : \; x_j > 0, \; j \in \{1, \ldots, n\}\}; \quad I^* = \phi.$$

Наконец, упомянем еще задачи однородного программирования:

$f = f_0$, $g_i = f_i + b$, $i \in I$; $b \in \mathbf{R}$, f_i, $i \in I \cup \{0\}$ – однородные функции степени s_i; $s_i \in \mathbf{R}$, $I^* = \phi$.

1.3. Классические задачи оптимизации

1.3.1. Предварительное замечание. Ниже мы кратко изложим необходимые условия для существования решений задачи

$$P_{KL}: \; \min\{f(x): \; x \in G\},$$

где $G = \{x \in \mathbf{R}^n : \; h_j(x) = 0, j \in I^*\}$, $I^* = \{1, \ldots, p\}$, $p < n$ (см. [37]).

Теорема 1.1. *Пусть функция* f: $D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$, *достигает в точке* $x^0 \in \operatorname{int} D(f)$ *локального минимума. Если существует градиент* $\nabla f(x^0)$, *то*

$$\nabla f(x^0) = 0. \tag{1.8}$$

Утверждение теоремы 1.1 в литературе часто называют принципом Ферма. Известные классические утверждения о существовании решений задачи P_{KL}, которые можно рассматривать как обобщения принципа Ферма, опираются на теорему о разрешимости системы уравнений.

1.3.2. Теорема о разрешимости системы уравнений. Система линейных уравнений

$$Ax = b, \tag{1.9}$$

где $A = (a^1, \ldots, a^n)$ – матрица порядка $p \times n$, $b \in \mathbf{R}^p$, $p \leqslant n$, в случае $r(A) = p$ может быть "разрешена" относительно p переменных. Полагая без ограничения общности $r(\bar{A}) = p$, где $\bar{A} = (a^1, \ldots, a^p)$, и обозначая $A = (\bar{A}, \bar{\bar{A}})$, $\bar{x} = (x_1, \ldots, x_p)'$, $\bar{\bar{x}} = (x_{p+1}, \ldots, x_n)'$, получим

$$\bar{x} = -\bar{A}^{-1}\bar{\bar{A}}\bar{\bar{x}} + \bar{A}^{-1}b. \tag{1.10}$$

При произвольном выборе x_i, $i = p+1, \ldots, n$, согласно (1.10) можно вычислить значения $x_1, \ldots, x_p$, которые вместе с x_i, $i = p+1, \ldots, n$, и дают решение системы (1.9).

Перейдем теперь к вопросу о разрешимости системы нелинейных уравнений

$$h_j(x) = 0, \quad j \in I^*. \tag{1.11}$$

В этом случае, вообще говоря, нельзя ожидать явного аналитического выражения для решений, аналогичного (1.10). При определенных предположениях могут быть, однако, сформулированы утверждения о локальной разрешимости, т.е. о разрешимости системы (1.11) в окрестности некоторой точки.

Пусть заданы непустое открытое множество $S \subseteq \mathbf{R}^{p+m}$ и p функций h_j: $S \to \mathbf{R}$, $j \in I^*$, непрерывных на S. Значение функции h_j в точке

$$(y^0, t^0)' = (y_1^0, \ldots, y_p^0, t_1^0, \ldots, t_m^0)' \in S$$

мы будем записывать кратко $h_j(y^0, t^0)$ вместо $h_j(y_1^0, \ldots, y_p^0, t_1^0, \ldots, t_m^0)$, $j \in I^*$. Соответственно вектор частных производных от h_j, $j \in I^*$, по $y_1, \ldots, y_p$ в точке $(y^0, t^0)' \in S$, если эти производные существуют, запишем в виде

$$\nabla_y h_j(y^0, t^0) = \left[\frac{\partial h_j(y^0, t^0)}{\partial y_1}, \ldots, \frac{\partial h_j(y^0, t^0)}{\partial y_p} \right]'.$$

Определитель, строки которого образованы векторами $\nabla_y h_j(y^0, t^0)$, $j \in I^*$, обозначим через

$$|\nabla_y h_{I^*}(y^0, t^0)'| = \left| \frac{\partial h_j(y^0, t^0)}{\partial y_k} \right|_{j,k \in I^*}.$$

Имеет место следующий результат.

Теорема 1.2 (*теорема о разрешимости*). *Пусть* $S \subseteq \mathbf{R}^{p+m}$ – *непустое открытое множество, функции* $h_j: S \to \mathbf{R}$, $j \in I^*$, *на множестве* S *непрерывны и имеют непрерывные частные производные по переменным* $y_1, \ldots, y_p$. *Предположим, что система*

$$h_j(y, t) = 0, \qquad j \in I^*, \tag{1.12}$$

имеет решение $(y^0, t^0)' \in S$ *и при этом*

$$|\nabla_y h_{I^*}(y^0, t^0)'| \neq 0. \tag{1.13}$$

Тогда в некоторой окрестности $T \subseteq \mathbf{R}^m$ *точки* t^0 *существует единственная система непрерывных функций* $\tilde{y}_j: T \to \mathbf{R}$, $j \in I^*$, *такая, что*

$$y^0 = \tilde{y}(t^0), \tag{1.14}$$

$$h_j(\tilde{y}(t), t) = 0, \tag{1.15}$$

где $\tilde{y}(t) = (\tilde{y}_1(t), \ldots, \tilde{y}_p(t))'$, $t = (t_1, \ldots, t_m)' \in T$.

Если, кроме того, h_j *обладает на* S *непрерывными частными производными по всем переменным* t_i, $i \in \{1, \ldots, m\}$, *то в некоторой окрестности точки* t^0 *существуют непрерывные первые частные производные* $\partial \tilde{y}_j / \partial t_i$, $(j, i) \in I^* \times \{1, \ldots, m\}$, *и выполняется равенство*

$$\sum_{k=1}^{p} \frac{\partial h_j(y, t)}{\partial y_k} \cdot \frac{\partial \tilde{y}_k(t)}{\partial t_i} + \frac{\partial h_j(y, t)}{\partial t_i} = 0, \qquad (j, i) \in I^* \times \{1, \ldots, m\}. \tag{1.16}$$

Доказательство этой известной теоремы классического анализа можно найти, например, в [28].

1.3.3. Множители Лагранжа. В этом разделе мы изложим классическую теорию множителей Лагранжа для задачи P_{KL} с ограничениями в форме равенств.

Пусть $D \subseteq \mathbf{R}^n$ – непустое открытое множество, $G \subseteq D$, и $f: D \to \mathbf{R}$, $h_j: D \to \mathbf{R}$ ($j \in I^*$) – дифференцируемые на D функции. Для произвольного фиксированного $x \in D$ определим матрицу порядка $p \times n$

$$H_0(x) = \nabla h_{I^*}(x)' = (\nabla h_1(x), \ldots, \nabla h_p(x))',$$

а также матрицу порядка $(p+1) \times n$

$$H(x) = \begin{bmatrix} \nabla h_{I^*}(x)' \\ \nabla f(x)' \end{bmatrix} = \begin{bmatrix} H_0(x) \\ \nabla f(x)' \end{bmatrix}.$$

Через $r(H_0(x))$ и $r(H(x))$ обозначим ранги этих матриц. С помощью этих обозначений сформулируем необходимые условия существования решений задачи P_{KL}.

Теорема 1.3. *Если* $x^0 \in G$ *и* $r(H(x^0)) = p + 1$, *то функция* f *не достигает в точке* x^0 *локального минимума на множестве* G.

Доказательство. Рассмотрим систему

$$f(x) - z = 0, \qquad h_j(x) = 0, \qquad j \in I^*, \tag{1.17}$$

из $p+1$ уравнений с $n+1$ неизвестными $x_1, \ldots, x_n, z$. По предположению существует невырожденная подматрица порядка $(p+1) \times (p+1)$ матрицы $H(x^0)$. Без ограничения общности можно считать, что эта подматрица

образована первыми $p+1$ столбцами $H(x^0)$. По теореме 1.2 в некоторой окрестности $T \subseteq \mathbf{R}^{n-p}$ точки $t^0 = (x^0_{p+2}, \ldots, x^0_n, f(x^0))'$ существуют однозначно определенные непрерывные функции $y_j, j \in \{1, \ldots, p+1\}$ такие, что

$$f(y_1(t), \ldots, y_{p+1}(t), \quad x_{p+2}, \ldots, x_n) - z = 0, \tag{1.18}$$

$$h_j(y_1(t), \ldots, y_{p+1}(t), \quad x_{p+2}, \ldots, x_n) = 0, \quad j \in I^*, \tag{1.19}$$

для всех

$$t = (x_{p+2}, \ldots, x_n, z) \in T. \tag{1.20}$$

Положим $y(t) = (y_1(t), \ldots, y_{p+1}(t))'$, $z_0 = f(x^0)$ и выберем $\delta^* < 0$ так, чтобы для всех $\delta \in (\delta^*, 0)$ выполнялось $t^\delta := \{x^0_{p+2}, \ldots, x^0_n, z_0 + \delta\} \in T$. Тогда для этих δ имеем

$$h_j(y(t^\delta), \quad x^0_{p+2}, \ldots, x^0_n) = 0, \quad j \in I^*,$$

$$f(y(t^\delta), \quad x^0_{p+2}, \ldots, x^0_n) = z_0 + \delta < z_0.$$

Поэтому в силу непрерывности функций y_j в любой (сколь угодно малой) окрестности точки $x^0 = (y(t^0), x^0_{p+2}, \ldots, x^0_n)$ найдутся точки x, для которых $f(x) < z_0$, так что функция f не может достигать в x^0 локального минимума на множестве G.

Для задачи P_{KL} введем *функцию Лагранжа*

$$L(x, u) = u_0 f(x) + \sum_{j=1}^{p} u_j h_j(x), \quad (x, u) \in D \times \mathbf{R}^{p+1}. \tag{1.21}$$

Назовем числа $u_0, u_1, \ldots, u_p$ *множителями Лагранжа*. Может быть доказан следующий критерий оптимальности для задачи P_{KL} *).

Теорема 1.4. *Если функция f достигает в $x^0 \in G$ локального минимума на множестве G, то $r(H(x^0)) \leqslant p$, т.е. существует нетривиальное решение $u^0 = (u^0_0, u^0_1, \ldots, u^0_p)' \in \mathbf{R}^{p+1}$ системы уравнений*

$$\nabla_x L(x^0, u) = u_0 \nabla f(x^0) + \sum_{j=1}^{p} u_j \nabla h_j(x^0) = 0. \tag{1.22}$$

Доказательство. Если $x^0 \in G$ – точка локального минимума f на G, то из теоремы 1.3 тотчас же следует, что $r(H(x^0)) \leqslant p$, а отсюда в силу условий разрешимости для систем линейных уравнений и вытекает соотношение (1.22).

Легко видеть, что (1.22) не является достаточным критерием оптимальности. Так, в случае линейной зависимости системы $\nabla h_j(x^0), j \in I^*$, всегда можно найти множитель Лагранжа u^0, у которого $u^0_0 = 0$, и при этом (1.22) выполняется независимо от роста f в окрестности точки x^0.

Из теоремы 1.4 при различных соотношениях между рангами матриц $H(x^0)$ и $H_0(x^0)$ вытекает следующее непосредственно проверяемое утверждение.

Теорема 1.5. *Пусть x^0 – точка локального минимума функции f на множестве G.*

а) *В случае $r(H(x^0)) > r(H_0(x^0))$ для любого нетривиального решения системы* (1.22) *выполняется $u^0_0 = 0$.*

---

*) См. в связи с этим § 4.2.

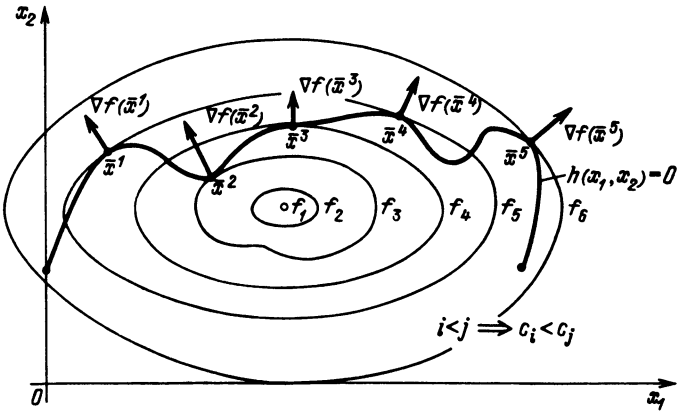


Рис. 1.1.

б) *В случае* $r(H(x^0)) = r(H_0(x^0))$ *существует нетривиальное решение системы* (1.22), *у которого* $u_0^0 \neq 0$.

Если, кроме того, $r(H(x^0)) = r(H_0(x^0)) = p$, *то, положив* $u_0^0 = 1$, *остальные множители Лагранжа* $u_1^0, \ldots, u_p^0$ *можно определить однозначно.*

З а м е ч а н и е. Утверждения теорем 1.3–1.5 имеют место и применительно к существованию локального максимума функции f в точке x^0 на множестве G.

При применении этих локальных критериев оптимальности поступают обычно следующим образом. Для функции Лагранжа задачи P_{KL} определяют частные производные по $x_1, \ldots, x_n$ и $u_1, \ldots, u_p$ и для полученной системы

$$u_0 \frac{\partial f(x)}{\partial x_i} + \sum_{j=1}^{p} u_j \frac{\partial h_j(x)}{\partial x_i} = 0, \quad i = 1, \ldots, n,$$

$$h_j(x) = 0, \quad j = 1, \ldots, p,$$

находят, если это практически возможно, все решения $(x_1, \ldots, x_n, u_0, u_1 \ldots \ldots, u_p)'$. Найденные при этом $\bar{x} = (\bar{x_1}, \ldots, \bar{x_n})' \in G$ являются теми точками, в которых функция f может достигать локального минимума или локального максимума на множестве G.

Тот факт, что не все найденные таким путем точки $\bar{x}$ являются локальными минимумами или максимумами, показывает представленный на рис. 1.1 пример с $n = 2$, $p = 1$, $G \subseteq \mathbf{R}^2$ и линиями уровня

$$f_i = \{(x_1, x_2) \in \mathbf{R}^2 : f(x_1, x_2) = c_i\}, \quad c_i < c_j \quad \text{для} \quad i < j.$$

Функция f достигает на множестве G в $\bar{x}^1, \bar{x}^4, \bar{x}^5$ локального максимума, в $\bar{x}^2$ – локального минимума, а в $\bar{x}^3$ не достигает ни локального максимума, ни локального минимума.

Во многих вопросах нелинейного программирования важную роль играет понятие выпуклости. Поэтому в данной главе мы изучим выпуклые множества, выпуклые функции и их обобщения, а также системы выпуклых неравенств. В основу нашего рассмотрения положено вещественное n-мерное евклидово векторное пространство $\mathbf{R}^n$.

2.1. Выпуклые множества и выпуклые конусы

2.1.1. К понятию выпуклого множества и выпуклого конуса. Множество $L \subseteq \mathbf{R}^n$ со свойством

$$x^1, x^2 \in L \Rightarrow \lambda_1 x^1 + \lambda_2 x^2 \in L \qquad \forall \lambda_1, \lambda_2 \in \mathbf{R} \tag{2.1а}$$

называется *подпространством.*

Множество $V \subseteq \mathbf{R}^n$ со свойством

$$x^1, x^2 \in V \Rightarrow \lambda_1 x^1 + \lambda_2 x^2 \in V \qquad \forall \lambda_1, \lambda_2 \in \mathbf{R}, \quad \lambda_1 + \lambda_2 = 1 \tag{2.1б}$$

называется *линейным многообразием.*

Множество $K \subseteq \mathbf{R}^n$ со свойством

$$x^1, x^2 \in K \Rightarrow \lambda_1 x^1 + \lambda_2 x^2 \in K \qquad \forall \lambda_1, \lambda_2 \in \mathbf{R}_+ \tag{2.1в}$$

называется *выпуклым конусом.*

Множество $C \subseteq \mathbf{R}^n$ со свойством

$$x^1, x^2 \in C \Rightarrow \lambda_1 x^1 + \lambda_2 x^2 \in C \qquad \forall \lambda_1, \lambda_2 \in \mathbf{R}_+, \qquad \lambda_1 + \lambda_2 = 1 \tag{2.1г}$$

называется *выпуклым множеством.*

Очевидно, что множества $\mathbf{R}^n$, $\{0\}$ и пустое множество отвечают каждому из четырех приведенных определений.

Далее, легко видеть, что подпространства, линейные многообразия и выпуклые конусы являются выпуклыми множествами. Любое подпространство является выпуклым конусом и линейным многообразием.

Пример 2.1. Пусть заданы произвольные фиксированные вектор $a \in \mathbf{R}^n$, $a \neq 0$, и число $\alpha \in \mathbf{R}$.

1. Гиперплоскость

$$E(a, \alpha) = \{x \in \mathbf{R}^n : a'x = \alpha\} \tag{2.2}$$

является линейным многообразием, а тем самым и выпуклым множеством.

2. Другие примеры выпуклых множеств представляют замкнутые полупространства

$$E_+(a, \alpha) = \{x \in \mathbf{R}^n : a'x \geqslant \alpha\},$$

$$E_-(a, \alpha) = \{x \in \mathbf{R}^n : a'x \leqslant \alpha\}$$

и открытые полупространства

$$\operatorname{int} E_+(a, \alpha) = \{x \in \mathbf{R}^n: \ a'x > \alpha\},$$

$$\operatorname{int} E_-(a, \alpha) = \{x \in \mathbf{R}^n: \ a'x < \alpha\}.$$

Гиперплоскость $E(a, \alpha)$ называется *образующей гиперплоскостью* для данных полупространств.

3. Пусть заданы матрица A порядка $m \times n$ и вектор $b \in \mathbf{R}^m$. Тогда множество

$$V = \{x: \ Ax = b\} \ *)$$

есть линейное многообразие и, следовательно, выпукло. Множество

$$V_+ = \{x: \ Ax \geqslant b\} \ **)$$

является пересечением m замкнутых полупространств

$$\{x \in \mathbf{R}^n: \ {}^i a x \geqslant b_i\}, \qquad i = 1, \ldots, m,$$

где ${}^i a$ – вектор-строка матрицы A.

С помощью определения легко проверить выпуклость V_+.

Вообще множество, которое может быть представлено как пересечение конечного числа замкнутых полупространств, будем называть *многогранным множеством.*

4. Если в примерах 2.1.1 – 2.1.3 положить $\alpha = 0$ и $b = 0$, то получим выпуклые конусы.

Множество $K_+ = \{x: \ Ax \geqslant 0\}$ является пересечением m ($m \geqslant 1$) замкнутых полупространств, образующие гиперплоскости которых есть подпространства. Такое множество назовем *многогранным конусом.* В противоположность произвольному многогранному множеству многогранный конус всегда не пуст.

В случае $A = E_n$ (E_n – единичная матрица порядка $n \times n$) получается неотрицательный ортант пространства $\mathbf{R}^n$: $\mathbf{R}^n_+ = \{x \in \mathbf{R}^n: \ x \geqslant 0\}$.

5. Открытый и замкнутый шары с центрами в точке x^0 и радиусами $\alpha > 0$:

$$S(x^0, \alpha) = \{x \in \mathbf{R}^n: \ \| x - x^0 \| < \alpha\}, \tag{2.3}$$

$$\bar{S}(x^0, \alpha) = \{x \in \mathbf{R}^n: \ \| x - x^0 \| \leqslant \alpha\} \tag{2.4}$$

являются выпуклыми множествами. Это легко может быть установлено с помощью неравенства треугольника.

Сегмент $[x^1, x^2]$, т.е. определенный двумя точками $x^1, x^2 \in \mathbf{R}^n$ отрезок, может быть записан в виде

$$[x^1, x^2] = \{x: \ x = \lambda_1 x^1 + \lambda_2 x^2; \quad \lambda_1, \lambda_2 \in [0, 1], \ \lambda_1 + \lambda_2 = 1\}. \tag{2.5}$$

Из определения выпуклого множества непосредственно следует, что множество $C \subseteq \mathbf{R}^n$ тогда и только тогда выпукло, когда вместе с любыми двумя точками $x^1, x^2 \in C$ оно содержит и сегмент $[x^1, x^2]$.

Пример 2.2 Очевидно, что множество $M_1 \subseteq \mathbf{R}^2$ на рис. 2.1 не выпукло. Также не выпукло множество

$$M_2 = \{x \in \mathbf{R}^2: \ x_1 \geqslant 0, \ x_2 \geqslant 0, \ x_1 + x_2 < 1\} \cup \left\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right\}$$

(рис. 2.2).

Напротив, множество M_3, полученное из M_2 удалением точки $x = (1, 0)'$, выпукло.

---

*) Линейное преобразование $A: \mathbf{R}^n \to \mathbf{R}^m$ и соответствующую матрицу A, для которых $y = Ax = A(x)$, $x \in \mathbf{R}^n$, мы не будем различать в обозначениях.

Образ множества $C \subseteq \mathbf{R}^n$ (при преобразовании A) обозначается через $A(C)$ или кратко AC. Прообраз множества $D \subseteq \mathbf{R}^n$ определяется как $A^- D = \{x \in \mathbf{R}^n: \ Ax \in D\}$.

**) Знаки $\leqslant$, $<$ между двумя векторами $x^1, x^2 \in \mathbf{R}^n$ следует понимать всегда в покоординатном смысле:

$$x^1 \leqslant x^2 \Rightarrow x^1_j \leqslant x^2_j, \qquad j = 1, \ldots, n,$$

$$x^1 < x^2 \Rightarrow x^1_j < x^2_j, \qquad j = 1, \ldots, n.$$

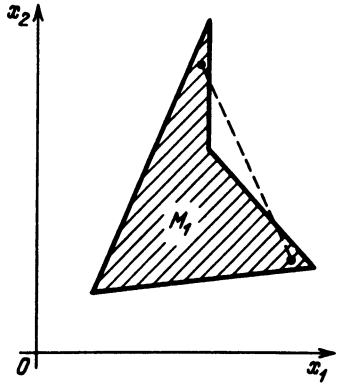


Рис. 2.1.

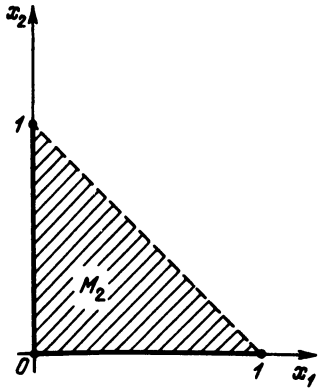


Рис. 2.2.

В дальнейшем большую роль играет понятие линейной комбинации векторов. В связи в этим мы введем сразу несколько конкретных видов линейной комбинации. Пусть заданы $m\,(m \geqslant 1)$ векторов $x^i \in \mathbf{R}^n$, $i \in I$, где $I = \{1, \ldots, m\}$. Выражение $\sum\limits_{i=1}^{m} \lambda_i x^i$ называется

а) при $(\lambda_1, \ldots, \lambda_m)' \in \mathbf{R}^m$ – *линейной комбинацией* векторов $x^i, i \in I$;

б) при $(\lambda_1, \ldots, \lambda_m)' \in \mathbf{R}^m_+$ – *неотрицательной линейной комбинацией* векторов $x^i, i \in I$;

в) при $(\lambda_1, \ldots, \lambda_m)' \in \mathbf{R}^m_+$, $(\lambda_1, \ldots, \lambda_m)' \neq 0$, – *положительной линейной комбинацией* векторов $x^i, i \in I$;

г) при $(\lambda_1, \ldots, \lambda_m)' \in P^m$, где $P^m = \{(\lambda_1, \ldots, \lambda_m)' \in \mathbf{R}^m: \sum\limits_{i=1}^{m} \lambda_i = 1\}$, – *аффинной линейной комбинацией векторов* $x^i, i \in I$;

д) при $(\lambda_1, \ldots, \lambda_m)' \in P^m_+$, где $P^m_+ = \mathbf{R}^m_+ \cap P^m$, – *выпуклой линейной комбинацией* векторов $x^i, i \in I$.

Векторы $x^i, i \in I, m \geqslant 1$, называются *линейно независимыми*, если однородное линейное уравнение

$$\sum_{i \in I} \lambda_i x^i = 0 \tag{2.6}$$

имеет лишь тривиальное решение $(\lambda_1, \ldots, \lambda_m)' = 0$. В противном случае векторы $x^i, i \in I$, *линейно зависимы*.

Векторы $x^i, i \in I, m \geqslant 2$, называются *аффинно независимыми*, если ни один из $x^j, j \in I$, не является аффинной линейной комбинацией остальных векторов $x^i, i \in I \setminus \{j\}$.

Это означает, что всякие $m-1$ векторов $x^i - x^j, i \in I \setminus \{j\}$ ($j \in I$ – произвольный фиксированный индекс), линейно независимы или, что то же самое, система

$$\sum_{i \in I} \lambda_i x^i = 0, \qquad \sum_{i \in I} \lambda_i = 0 \tag{2.7}$$

имеет лишь тривиальное решение. В противном случае векторы x^i, $i \in I$, называются *аффинно зависимыми*.

Теперь мы можем охарактеризовать выпуклое множество, линейное многообразие и т.д. следующим образом.

Т е о р е м а 2.1. *Множество* $M \subseteq \mathbf{R}^n$ *тогда и только тогда является*:

а) *выпуклым множеством, когда*

$$x^1, \ldots, x^m \in M \Rightarrow \sum_{i=1}^{m} \lambda_i x^i \in M \quad \forall (\lambda_1, \ldots, \lambda_m)' \in P_+^m; \tag{2.8а}$$

б) *линейным многообразием, когда*

$$x^1, \ldots, x^m \in M \Rightarrow \sum_{i=1}^{m} \lambda_i x^i \in M \quad \forall (\lambda_1, \ldots, \lambda_m)' \in P^m; \tag{2.8б}$$

в) *выпуклым конусом, когда*

$$x_1, \ldots, x^m \in M \Rightarrow \sum_{i=1}^{m} \lambda_i x^i \in M \quad \forall (\lambda_1, \ldots, \lambda_m)' \in \mathbf{R}_+^m; \tag{2.8в}$$

г) *подпространством, когда*

$$x^1, \ldots, x^m \in M \Rightarrow \sum_{i=1}^{m} \lambda_i x^i \in M \quad \forall (\lambda_1, \ldots, \lambda_m)' \in \mathbf{R}^m; \tag{2.8г}$$

где m *– произвольное натуральное число.*

Д о к а з а т е л ь с т в о. Мы ограничимся лишь проверкой справедливости первого утверждения.

1. Достаточность условия (2.8а). Полагая в (2.8а) $m = 2$, непосредственно получаем, что M выпукло.

2. Необходимость условия (2.8а). Пусть M выпукло. Тогда (2.8а) выполняется при $m = 2$. Предположим, что (2.8а) выполнено для $m = k$ $(k \geqslant 2)$, и покажем, что оно выполняется и при $m = k + 1$. Рассмотрим для этого

$$x^1, \ldots, x^k, x^{k+1} \in M, \qquad (\lambda_1, \ldots, \lambda_{k+1}) \in P_+^{k+1}.$$

Если $\lambda_{k+1} = 1$, то в силу $\sum_{i=1}^{k+1} \lambda_i x^i = x^{k+1} \in M$ требуемое утверждение доказано.

В случае $\lambda_{k+1} < 1$ имеем $\lambda_0 = \sum_{i=1}^{k} \lambda_i \neq 0$; поэтому

$$\sum_{i=1}^{k+1} \lambda_i x^i = \lambda_0 \sum_{i=1}^{k} \frac{\lambda_i}{\lambda_0} x^i + \lambda_{k+1} x^{k+1}. \tag{2.9}$$

Так как $\frac{\lambda_1}{\lambda_0}, \ldots, \frac{\lambda_k}{\lambda_0} \in [0, 1]$ и $\sum_{i=1}^{k} \frac{\lambda_i}{\lambda_0} = 1$, то по предположению индукции $x^0 = \sum_{i=1}^{k} \frac{\lambda_i}{\lambda_0} x^i \in M$. Отсюда по определению выпуклого множества $\sum_{i=1}^{k+1} \lambda_i x^i = \lambda_0 x^0 + \lambda_{k+1} x^{k+1} \in M$, поскольку $\lambda_0, \lambda_{k+1} \in [0, 1]$; $\lambda_0 + \lambda_{k+1} = \sum_{i=1}^{k} \lambda_i + \lambda_{k+1} = 1$.

Теорема 2.1 показывает, что M тогда и только тогда является выпуклым множеством (линейным многообразием, выпуклым конусом, подпространством), когда любая выпуклая линейная комбинация (аффинная линейная комбинация, неотрицательная линейная комбинация, линейная комбинация) m точек из M принадлежит M.

2.1.2. Операции над выпуклыми множествами. Пусть $M_1, M_2 \subseteq \mathbf{R}^n$ – произвольные множества. Множество

$$M_1 + M_2 = \{z:\ z = x + y, \quad x \in M_1, \quad y \in M_2\} \tag{2.10}$$

называется *суммой* M_1 и M_2, а множество

$$\lambda M_1 = \{z:\ z = \lambda x, \quad x \in M_1\} \tag{2.11}$$

– *скалярным кратным множества* M_1 со скаляром λ.

Для заданных матрицы A порядка $m \times n$ и вектора $b \in \mathbf{R}^m$ с помощью соотношения

$$y = b + Ax, \quad x \in M_1,$$

определяется *аффинное преобразование* множества M_1 в $\mathbf{R}^m$. При $b = 0$ получаем *линейное преобразование*

$$y = Ax, \quad x \in M_1.$$

Мы условимся, что

$$M_1 + \phi = \phi, \quad -M_1 = (-1)M_1, \quad M_1 - M_2 = M_1 + (-M_2). \tag{2.12}$$

Если M_2 – одноэлементное множество $\{a\}$, то вместо $M_1 + \{a\}$ мы будем просто писать $M_1 + a$ и говорить о *сдвиге* M_1 на вектор a. Заметим, что $\lambda M_1 + a$ получается из M_1 при $\lambda > 0$ с помощью собственного и при $\lambda < 0$ с помощью несобственного преобразований подобия.

Можно легко установить следующие свойства множеств:

$$\begin{gathered} M_1 + \{0\} = M_1, \quad M_1 + M_2 = M_2 + M_1, \\ \lambda(M_1 + M_2) = \lambda M_1 + \lambda M_2, \quad (\lambda_1 + \lambda_2)M_1 \subseteq \lambda_1 M_1 + \lambda_2 M_1. \end{gathered} \tag{2.13}$$

Равенство $2M = M + M$, вообще говоря, не имеет места, а справедливо лишь включение $2M \subseteq M + M$. Далее, справедливо $M - M \supseteq \{0\}$ (и, вообще говоря, не выполняется $M - M = \{0\}$).

Пример 2.3. На рис. 2.3 для множеств $M_1, M_2 \subseteq \mathbf{R}^2$ построена сумма $M_1 + M_2$.

Каждое линейное многообразие V можно представить в виде

$$V = a + L, \tag{2.14}$$

где a – произвольная точка из V и L – некоторое подпространство в $\mathbf{R}^n$. В этом случае говорят что L параллельно многообразию V. Параллельное многообразию V подпространство L получается из соотношения

$$V - V = L$$

(см. рис. 2.4).

Обратно, для произвольного подпространства $L \subseteq \mathbf{R}^n$ и $a \in \mathbf{R}^n$ множество $a + L$ является линейным многообразием.

Связь между суммой и пересечением двух множеств устанавливается в следующем утверждении.

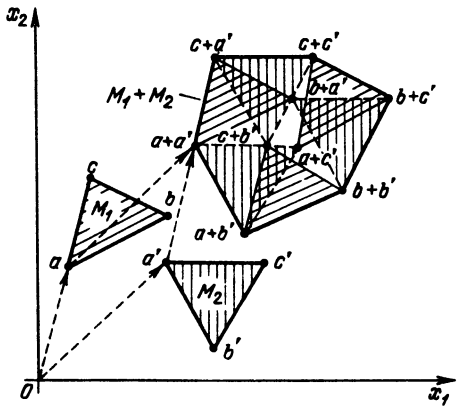


Рис. 2.3.

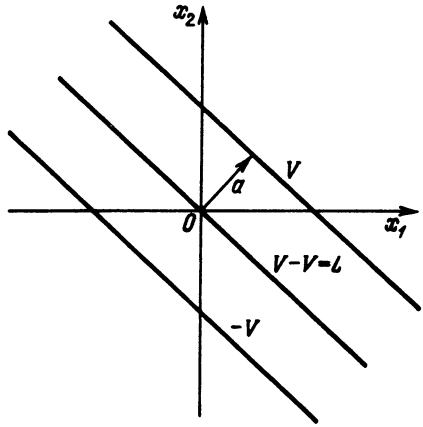


Рис. 2.4.

Т е о р е м а 2.2. *Для заданных множеств* $M_1, M_2, M_3 \subseteq \mathbf{R}^n$ *справедливо соотношение*

$$\{M_1 + M_2\} \cap M_3 = \phi \Longleftrightarrow M_1 \cap (M_3 - M_2) = \phi.$$

Д о к а з а т е л ь с т в о. Из непустоты $M_1 \cap (M_3 - M_2)$ следует существование $x^1 \in M_1$ такого, что $x^1 = x^3 - x^2$, где $x^2 \in M_2$, $x^3 \in M_3$. Отсюда $x^3 = x^1 + x^2$ и, следовательно, $(M_1 + M_2) \cap M_3 \neq \phi$.

Тем самым показано, что

$$(M_1 + M_2) \cap M_2 = \phi \Rightarrow M_1 \cap (M_3 - M_2) = \phi.$$

Если здесь заменить M_2 на $-M_2$, M_1 на M_3 и M_3 на M_1, то получим

$$(M_3 - M_2) \cap M_1 = \phi \Rightarrow M_3 \cap (M_1 + M_2) = \phi,$$

что и доказывает теорему.

Теперь рассмотрим множество $\lambda_1 M + \lambda_2 M$ для конкретных скаляров λ_1, λ_2 и конкретных множеств M.

Имеют место следующие утверждения:

1. C – выпуклое множество, $(\lambda_1, \lambda_2)' \in \mathbf{R}^2_+$ $\Rightarrow \lambda_1 C + \lambda_2 C = (\lambda_1 + \lambda_2) C.$ (2.15)

Отсюда, в частности, следует

C – выпуклое множество, $(\lambda_1, \lambda_2)' \in P^2_+$ $\Rightarrow \lambda_1 C + \lambda_2 C = C.$ (2.16)

2. K – выпуклый конус, $(\lambda_1, \lambda_2)' \in \mathbf{R}^2_+ \setminus \{0\}$ $\Rightarrow \lambda_1 K + \lambda_2 K = K.$ (2.17)

3. V – линейное многообразие, $(\lambda_1, \lambda_2)' \in P^2$ $\Rightarrow \lambda_1 V + \lambda_2 V = V.$ (2.18)

4. L – подпространство, $(\lambda_1, \lambda_2)' \in \mathbf{R}^2 \setminus \{0\}$ $\Rightarrow \lambda_1 L + \lambda_2 L = L.$ (2.19)

Т е о р е м а 2.3. *Пересечение произвольного числа выпуклых множеств (выпуклых конусов, линейных многообразий, подпространств) есть выпуклое множество (выпуклый конус, линейное многообразие, подпространство).*

Д о к а з а т е л ь с т в о. Ограничимся случаем выпуклых множеств.

Пусть $(C_i)_{i \in Z}$ – семейство выпуклых множеств (Z – произвольное множество индексов). Далее, пусть $x,\ y \in \bigcap_{i \in Z} C_i$ и $(\lambda_1, \lambda_2)' \in P_+^2$. Тогда $x, y \in C_i\ \forall i \in Z$. Поскольку C_i выпуклы, то $\lambda_1 x + \lambda_2 y \in C_i\ \forall i \in Z$, что и доказывает утверждение.

Т е о р е м а 2.4. 1. *Пусть $C_1, C_2 \subseteq \mathbf{R}^n$ – выпуклые множества, $\lambda \in \mathbf{R}$ – произвольный скаляр, A: $\mathbf{R}^n \to \mathbf{R}^m$ – линейное преобразование. Тогда множества $C_1 + C_2$, λC_1, AC_1 выпуклы.*

2. *Пусть $C_1 \subseteq \mathbf{R}^n$, $C_2 \subseteq \mathbf{R}^q$ – выпуклые множества. Тогда декартово произведение $C_1 \times C_2$ выпукло.*

3. *Аналогичные утверждения справедливы для выпуклых конусов, линейных многообразий и подпространств.*

Д о к а з а т е л ь с т в о. Мы приведем доказательство для выпуклых множеств.

1. а) Имеем

$$\left.\begin{array}{l} z^1 = x^1 + y^1, \quad z^2 = x^2 + y^2 \in C_1 + C_2, \\ x^1,\ x^2 \in C_1, \quad y^1, y^2 \in C_2,\ (\lambda_1, \lambda_2)' \in P_+^2 \end{array}\right\} \Rightarrow \left\{\begin{array}{l} \lambda_1 x^1 + \lambda_2 x^2 \in C_1, \\ \lambda_1 y^1 + \lambda_2 y^2 \in C_2. \end{array}\right.$$

Отсюда

$$z = \lambda_1 z^1 + \lambda_2 z^2 = (\lambda_1 x^1 + \lambda_2 x^2) + (\lambda_1 y^1 + \lambda_2 y^2) \in C_1 + C_2.$$

б) Имеем

$$\left.\begin{array}{l} z^1 = \lambda x^1, \quad z^2 = \lambda x^2 \in \lambda C_1, \\ x^1, x^2 \in C_1, \quad (\lambda_1, \lambda_2) \in P_+^2 \end{array}\right\} \Rightarrow \lambda_1 x^1 + \lambda_2 x^2 \in C_1,$$

и отсюда

$$z = \lambda_1 z^1 + \lambda_2 z^2 = \lambda(\lambda_1 x^1 + \lambda_2 x^2) \in \lambda C_1.$$

в) Имеем

$$\left.\begin{array}{l} y^1, y^2 \in AC_1, \\ y^1 = Ax^1,\ x^1 \in C_1,\ y_2 = Ax^2, \\ x^1 \in C_1;\ \ (\lambda_1, \lambda_2)' \in P_+^2 \end{array}\right\} \Rightarrow \left\{\begin{array}{l} y = \lambda_1 y^1 + \lambda_2 y^2 = \lambda_1 Ax^1 + \lambda_2 Ax^2 = \\ = A(\lambda_1 x^1 + \lambda_2 x^2) \in AC_1. \end{array}\right.$$

2. Имеем

$$\left.\begin{array}{l} z^1, z^2 \in C_1 \times C_2, \\ z^1 = (x^1, y^1)', \quad z^2 = (x^2, y^2)', \\ x^1, x^2 \in C_1, \quad y^1, y^2 \in C_2;\ (\lambda_1, \lambda_2)' \in P_+^2 \end{array}\right\} \Rightarrow \left\{\begin{array}{l} \lambda_1 x^1 + \lambda_2 x^2 \in C_1, \\ \lambda_1 y^1 + \lambda_2 y^2 \in C_2, \end{array}\right.$$

и поэтому

$$z = \lambda_1 z^1 + \lambda_2 z^2 = \lambda_1 \begin{pmatrix} x^1 \\ y^1 \end{pmatrix} + \lambda_2 \begin{pmatrix} x^2 \\ y^2 \end{pmatrix} \in C_1 \times C_2.$$

Т е о р е м а 2.5. *Пусть* $C \subseteq \mathbf{R}^n$ *– выпуклое множество и* T: $\mathbf{R}^n \to \mathbf{R}^m$ *– аффинное преобразование. Тогда* TC *– тоже выпуклое множество.*

Д о к а з а т е л ь с т в о. Аффинное преобразование может быть рассмотрено как результат последовательного применения линейного преобразования и сдвига.

Утверждение, аналогичное теореме 2.5, справедливо и для линейных многообразий.

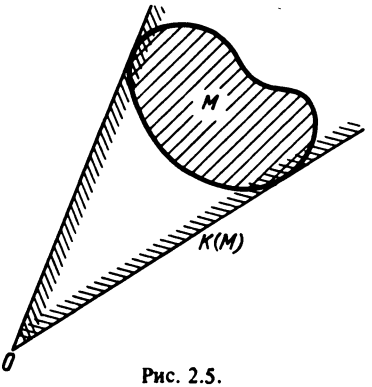


Рис. 2.5.

Введем понятие конуса как обобщение понятия выпуклого конуса.

Множество $K \subseteq \mathbf{R}^n$ называется *конусом*, если справедливо соотношение

$$x \in K, \quad \lambda \geqslant 0 \Rightarrow \lambda x \in K. \tag{2.20}$$

Соотношение (2.20) равносильно тому, что

$$\lambda K \subseteq K \quad \forall \lambda \geqslant 0. \tag{2.21}$$

Пусть K – произвольный конус. Множество $-K$, которое, очевидно, тоже является конусом, называется *противоположным конусом* для K.

Любой выпуклый конус является конусом в смысле данного выше определения. Чтобы показать это, достаточно положить в (2.1в) $x^1 = x^2$.

Произвольное множество $M \subseteq \mathbf{R}^n$ порождает конус

$$K(M) = \{ x \in \mathbf{R}^n: \quad x = \lambda y, \quad \lambda \geqslant 0, \ y \in M \} \tag{2.22}$$

(см. рис. 2.5).

Если, в частности, $M = \{a\}$, $a \neq 0$, то порожденный вектором a конус $K(\{a\}) = K(a)$ называется *полупрямой* или *лучом:*

$$K(a) = \{ x \in \mathbf{R}^n: \quad x = \lambda a, \ \lambda \geqslant 0 \}. \tag{2.23}$$

Луч $K(a)$ является выпуклым конусом.

Далее, справедливо следующее утверждение.

Т е о р е м а 2.6. *Если* $C \subseteq \mathbf{R}^n$ *– выпуклое множество, то* $K(C)$ *– выпуклый конус.*

Д о к а з а т е л ь с т в о. Так как согласно (2.22) $K(C)$ – конус, то остается показать, что $K(C)$ – выпуклое множество.

По предположению для $x^1, x^2 \in K(C)$ существуют точки $y^1, y^2 \in C$ и числа $\lambda_1, \lambda_2 \in \mathbf{R}_+$ такие, что $x^1 = \lambda_1 y^1$, $x^2 = \lambda_2 y^2$. Для произвольного $(\mu_1, \mu_2)' \in P_+^2$ образуем

$$x = \mu_1 x^1 + \mu_2 x^2 = \mu_1 \lambda_1 y^1 + \mu_2 \lambda_2 y^2 \tag{2.24}$$

и рассмотрим два случая.

С л у ч а й 1. $\mu_1 \lambda_1 + \mu_2 \lambda_2 > 0$.

Полагая

$$\xi_1 = \frac{\mu_1 \lambda_1}{\mu_1 \lambda_1 + \mu_2 \lambda_2}, \quad \xi_2 = 1 - \frac{\mu_1 \lambda_1}{\mu_1 \lambda_1 + \mu_2 \lambda_2}, \tag{2.25}$$

получим $(\xi_1, \xi_2)' \in P_+^2$. Поэтому из (2.24) с учетом выпуклости C вытекает

$x = (\mu_1\lambda_1 + \mu_2\lambda_2)(\xi_1 y^1 + \xi_2 y^2) \in K(C)$.

Случай 2. $\mu_1\lambda_1 + \mu_2\lambda_2 = 0$.

Отсюда следует $\mu_1\lambda_1 = \mu_2\lambda_2 = 0$ и, следовательно, $\{x = 0\} \in K(C)$.

Пример 2.4. Выпуклое множество $C = P_+^n$ порождает выпуклый конус $K(C) = \mathbf{R}_+^n$.

Для установления связи между произвольным множеством $M \subseteq \mathbf{R}^n$ и множествами C, V, K и L из (2.1а) – (2.1г) введем ряд понятий.

Отображение $H: P(\mathbf{R}^n) \to P(\mathbf{R}^n)$ множества $P(\mathbf{R}^n)$ всех подмножеств $\mathbf{R}^n$ в себя называется *операцией взятия оболочки*, если оно обладает следующими свойствами:

$$1.\ M \subseteq H(M);$$

$$2.\ M_1 \subseteq M_2 \Rightarrow H(M_1) \subseteq H(M_2); \tag{2.26}$$

$$3.\ H(H(M)) = H(M).$$

Можно показать, что следующие пять понятий являются операциями взятия оболочки в смысле данного выше определения. Пусть всюду в дальнейшем Z – соответствующее множество индексов.

Выпуклой оболочкой $C(M)$ множества $M \subseteq \mathbf{R}^n$ называется пересечение всех выпуклых множеств C_i, содержащих M:

$$C(M) = \bigcap_{i \in Z} C_i, \quad \text{где } C_i \supseteq M. \tag{2.27}$$

Конической оболочкой $K(M)$ множества $M \subseteq \mathbf{R}^n$ называется пересечение всех конусов K_i, содержащих M:

$$K(M) = \bigcap_{i \in Z} K_i, \quad \text{где } K_i \supseteq M. \tag{2.28}$$

Выпуклой конической оболочкой или *положительной оболочкой* $K^c(M)$ множества $M \subseteq \mathbf{R}^n$ называется пересечение всех выпуклых конусов K_i^c, содержащих M:

$$K^c(M) = \bigcap_{i \in Z} K_i^c, \quad \text{где } K_i^c \supseteq M. \tag{2.29}$$

Аффинной оболочкой $V(M)$ множества $M \subseteq \mathbf{R}^n$ называется пересечение всех линейных многообразий V_i, содержащих M:

$$V(M) = \bigcap_{i \in Z} V_i, \quad \text{где } V_i \supseteq M. \tag{2.30}$$

Линейной оболочкой $L(M)$ множества $M \subseteq \mathbf{R}^n$ называется пересечение всех подпространств L_i, содержащих M:

$$L(M) = \bigcap_{i \in Z} L_i, \quad \text{где } L_i \supseteq M. \tag{2.31}$$

Из введенных выше определений оболочек вытекает, в частности, что оболочка пустого множества есть пустое множество.

Из определений непосредственно следует, что, например, $C(M)$ есть наименьшее выпуклое множество, содержащее M. Аналогичные утверждения справедливы для $K(M)$, $K^c(M)$, $V(M)$ и $L(M)$.

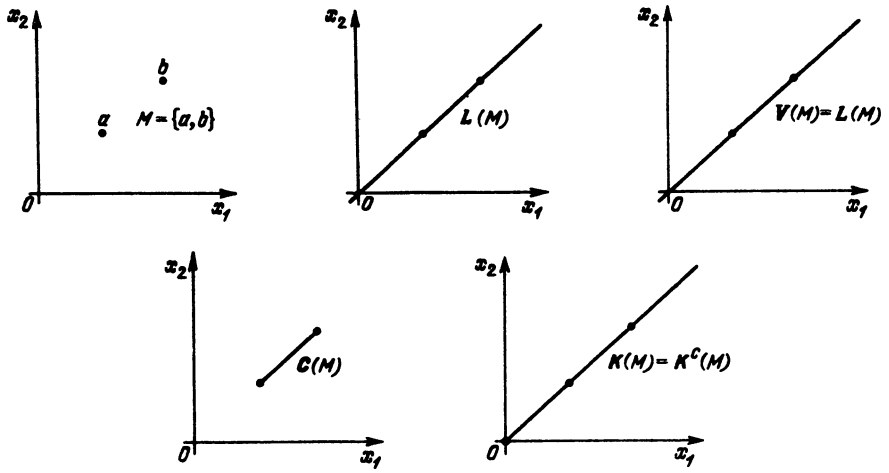


Рис. 2.6.

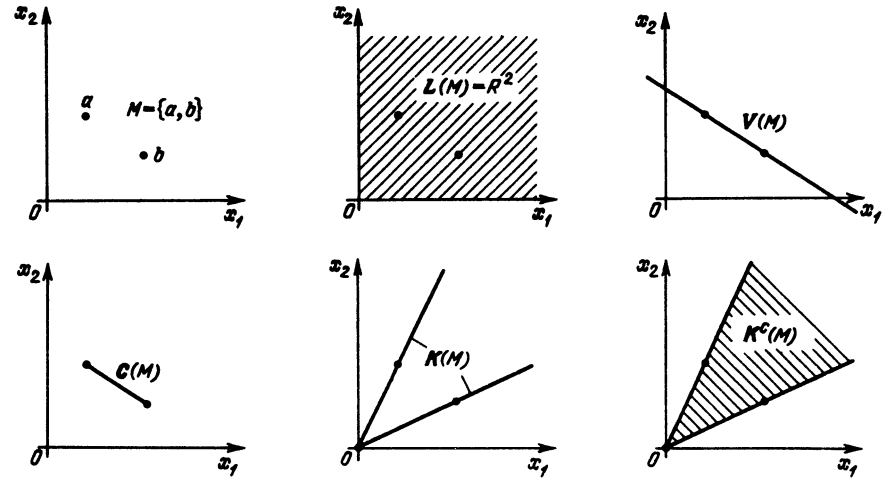


Рис. 2.7.

В целях наглядности мы изобразим на рис. 2.6, 2.7 для двухэлементных множеств соответствующие оболочки.

Пример 2.5. Для множества $M=\left\{x=(x_1,\ x_2)'\in\mathbf{R}^2:\ 0\leqslant x_2\leqslant\dfrac{1}{1+x_1^2}\right\}$ имеем $C(M)=\{x\in\mathbf{R}^2:\ 0\leqslant x_2<1\}\cup\{(0,1)'\}$.

Пример 2.6. Если $M=\left\{x\in\mathbf{R}:\ x=\dfrac{1}{k},\ k\in\mathbf{N}\right\}$, то $C(M)=(0,1]$.

Справедливо следующее утверждение о представлении введенных выше оболочек.

Т е о р е м а 2.7. *Пусть* $M \subseteq \mathbf{R}^n$ *– произвольное множество. Тогда*:

$$1.\ C(M) = \{ x: x = \sum_{i \in J} \lambda_i x^i,\quad \lambda_i \geqslant 0,\ \sum_{i \in J} \lambda_i = 1;$$

$$x^i \in M,\ J \subset \mathbf{R},\ J \text{ конечно} \}, \tag{2.32}$$

т.е. $C(M)$ *совпадает с множеством всех выпуклых линейных комбинаций точек из* M;

$$2.\ K^c(M) = \{ x:\ x = \sum_{i \in J} \lambda_i x^i,\ \lambda_i \geqslant 0,\ x^i \in M,\ J \subset \mathbf{R},\ J \text{ конечно} \}, \tag{2.33}$$

т.е. $K^c(M)$ *совпадает с множеством всех неотрицательных линейных комбинаций точек из* M;

$$3.\ V(M) = \{ x:\ x = \sum_{i \in J} \lambda_i x^i,\ \sum_{i \in J} \lambda_i = 1,\ x^i \in M,\ J \subset \mathbf{R},\ J \text{ конечно} \}, \tag{2.34}$$

т.е. $V(M)$ *совпадает с множеством всех аффинных линейных комбинаций точек из* M;

$$4.\ L(M) = \{ x:\ x = \sum_{i \in J} \lambda_i x^i,\ \lambda_i \in \mathbf{R},\ x^i \in M,\ J \subset \mathbf{R}, J \text{ конечно} \}, \tag{2.35}$$

т.е. $L(M)$ *совпадает с множеством всех линейных комбинаций точек из* M;

$$5.\ K(M) = K(M), \tag{2.36}$$

т.е. коническая оболочка $K(M)$ *совпадает с конусом, порожденным множеством* M.

Д о к а з а т е л ь с т в о. Мы ограничимся проверкой справедливости первого утверждения; доказательство остальных может быть проведено аналогичным образом.

Пусть $C(M)$ – множество всех выпуклых линейных комбинаций точек из M. Если C_l – выпуклое множество и $C_l \supseteq M$, то в силу теоремы 2.1 $C_l \supseteq C(M)$. Из (2.27) следует

$$C(M) \supseteq C(M). \tag{2.37}$$

Обратно, любая выпуклая линейная комбинация точек, которые в свою очередь являются выпуклыми линейными комбинациями точек из M, – также выпуклая линейная комбинация точек из M. А именно, для $\rho \in [0, 1)$ выполняется

$$\rho \sum_{i \in J_1} \lambda_i x^i + (1-\rho) \sum_{i \in J_2} \overline{\lambda}_i \overline{x}^i = \sum_{i \in J_1} (\rho \lambda_i) x^i + \sum_{i \in J_2} ([1-\rho]\,\overline{\lambda}_i)\,\overline{x}^i,$$

и из $\sum_{i \in J_1} \lambda_i = 1$, $\sum_{i \in J_2} \overline{\lambda}_i = 1$, $\lambda_i, \overline{\lambda}_i \geqslant 0$ следует

$$\sum_{i \in J_1} (\rho \lambda_i) + \sum_{i \in J_2} ([1-\rho])\,\overline{\lambda}_i = 1.$$

Поэтому

$$C(M) \subseteq C(M). \tag{2.38}$$

Из (2.37) и (2.38) получаем $C(M) = C(M)$.

Выпуклая оболочка конечного множества $M = \{x^1, \ldots, x^{m+1}\} \subseteq \mathbf{R}^n$ называется *выпуклым многогранником* (*политопом*). Если точки $x^1, \ldots, x^{m+1}$ аффинно независимы, то $C(M)$ называется *m-мерным симплексом* S^m с вершинами $x^1, \ldots, x^{m+1}$.

Таким образом, в силу утверждения 1 теоремы 2.7 выпуклая оболочка $C(M)$ произвольного множества $M \subseteq \mathbf{R}^n$ является объединением всех выпуклых многогранников, которые образованы точками из M. Это утверждение можно усилить, показав, что $C(M)$ в случае $\dim M = m$ есть объединение всех m-мерных симплексов, вершины которых лежат в M *).

Т е о р е м а 2.8 (*теорема Каратеодори для конусов*). *Пусть* $M \subseteq \mathbf{R}^n$ – *множество такое, что* $\dim K^c(M) = m$. *Тогда*:

1. *Любая точка* $x \in K^c(M)$ *может быть представлена в виде неотрицательной линейной комбинации не более чем m линейно независимых векторов из M.*

2. $K^c(M) = \bigcup_{i \in Z} K^c(M_i)$, где $(M_i)_{i \in Z}$ – *семейство всех множеств* M_i, $i \in Z$, *которые состоят из m линейно независимых векторов множества M.*

Д о к а з а т е л ь с т в о. 1. Соотношение $x \in K^c(M)$ означает, что

$$x = \sum_{i=1}^{p} \lambda_i a^i, \quad a^i \in M_i, \quad i = 1, \ldots, p, \quad (\lambda_1, \ldots, \lambda_p)' \in \mathbf{R}_+^p. \tag{2.39}$$

Допустим, что $p > m$ и $\lambda_i > 0$ для $i = 1, \ldots, p$. Так как $\dim K^c(M) = m$, то векторы $a^1, \ldots, a^p$ линейно зависимы, т.е. уравнение

$$\sum_{i=1}^{p} v_i a^i = 0 \tag{2.40}$$

имеет решение $(v_1, \ldots, v_p)' \neq 0$.

Для некоторого нетривиального решения, имеющего без ограничения общности по меньшей мере одну положительную координату, положим $\mu = \max v_i \lambda_i^{-1}$. Отсюда

$$\lambda_i - \frac{v_i}{\mu} \geqslant 0, \quad i = 1, \ldots, p, \quad x = \sum_{i=1}^{p} \left(\lambda_i - \frac{v_i}{\mu}\right) a^i, \tag{2.41}$$

при этом хотя бы один коэффициент в (2.41) равен нулю. Тем самым число тех a^i, которые входят в положительную линейную комбинацию (2.41) с отличными от нуля коэффициентами, сократилось по сравнению с (2.39) по крайней мере на единицу. Эта редукция может быть повторена, пока не получим $p \leqslant m$.

2. Мы можем считать, что вектор $x \in K^c(M)$ представим в виде неотрицательной линейной комбинации ровно m линейно независимых векторов из M: в случае, если в (2.39) $p < m$, всегда можно расширить неотрицательную линейную комбинацию таким образом, чтобы коэффициенты λ_i при

---

*) Здесь $\dim M$ – *размерность множества* $M \subseteq \mathbf{R}^n$ – определяется как $\dim M = \dim V(M)$. Для $V(M) = \phi$ мы полагаем $\dim M = -1$. *Размерность линейного многообразия* V определяется как размерность параллельного V подпространства L.

вновь вошедших векторах $a^i \in M$ равнялись нулю и одновременно a^i, $i = 1, \ldots, m$, были линейно независимы. Тем самым каждая точка из $K^c(M)$ принадлежит множеству $K^c(M_l)$, $i \in Z$, где $(M_l)_{i \in Z}$ – семейство всех множеств, состоящих из m линейно независимых векторов множества M. Очевидно, тогда

$$K^c(M) \subseteq \bigcup_{i \in Z} K^c(M_l).$$

Из определения множеств M_i, $i \in Z$, и из второго свойства операции взятия оболочки в (2.26) вытекает, что $K^c(M) \supseteq K^c(M_i)$, $i \in Z$, откуда

$$K^c(M) \supseteq \bigcup_{i \in Z} K^c(M_i).$$

Следовательно,

$$K^c(M) = \bigcup_{i \in Z} K^c(M_i).$$

Теорема доказана полностью.

Из теоремы 2.8 можно теперь вывести уже упоминавшееся утверждение для выпуклых множеств.

Т е о р е м а 2.9 (*теорема Каратеодори*). *Пусть* $M \subseteq \mathbf{R}^n$ – *множество, такое, что* $\dim M = m$. *Тогда*:

1. *Любая точка* $x \in C(M)$ *может быть представлена как выпуклая линейная комбинация не более чем* $m + 1$ *векторов из* M.

2. $C(M) = \bigcup_{i \in Z} S_i^m$, *где* $(S_i^m)_{i \in Z}$ – *семейство* m-*мерных симплексов вида* $S_i^m = C(M_l)$, $i \in Z$, *а каждое* M_l, $i \in Z$, *состоит из* $m + 1$ *аффинно независимых векторов множества* M.

Д о к а з а т е л ь с т в о. Поскольку $\dim M = m$, в M существует не более $m + 1$ аффинно независимых векторов. Это равносильно существованию не более чем $m + 1$ векторов $a^i \in M$, $i = 1, \ldots, m + 1$, таких, что уравнение

$$\sum_{i=1}^{m+1} \begin{pmatrix} a^i \\ 1 \end{pmatrix} \mu_i = 0, \quad \begin{pmatrix} a^i \\ 1 \end{pmatrix} \in \mathbf{R}^{n+1}, \quad i = 1, \ldots, m + 1,$$

имеет лишь тривиальное решение $(\mu_1, \ldots, \mu_{m+1})' = 0$. Последнее означает, что

$$\dim L(\tilde{M}) = \dim K^c(\tilde{M}) = m + 1,$$

где $\tilde{M} = \left\{ \tilde{x} \in \mathbf{R}^{n+1} : \tilde{x} = \begin{pmatrix} x \\ 1 \end{pmatrix}, \ x \in M \right\}$.

Легко сообразить, что $\tilde{x} \in K^c(M)$ тогда и только тогда, когда $x \in C(M)$ (рис. 2.8).

Применяя к множеству $\tilde{M}$ теорему 2.8, получим с учетом равенства $\dim K^c(\tilde{M}) = m + 1$, что любая точка из $K^c(\tilde{M})$ может быть представлена в виде неотрицательной линейной комбинации не более чем $m + 1$ линейно независимых векторов из $\tilde{M}$; соответственно любая точка из $C(M)$ может быть представлена в виде выпуклой линейной комбинации не более чем $m + 1$ точек из M.

Точно так же из теоремы 2.8 вытекает второе утверждение теоремы 2.9. Вспомним, что выпуклая оболочка $m + 1$ аффинно независимых точек есть

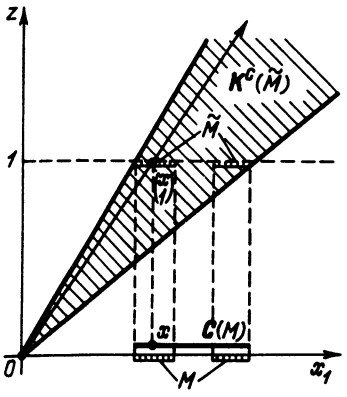


Рис. 2.8.

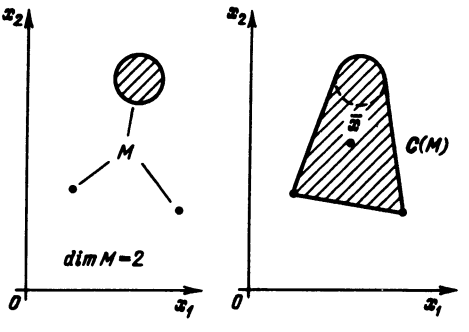


Рис. 2.9.

m-мерный симплекс S^m. Таким образом, справедливы следующие импликации

$$x \in C(M) \Longleftrightarrow \tilde{x} \in K^c(\tilde{M}) \Longleftrightarrow \tilde{x} \in \bigcup_{i \in Z} K^c(\tilde{M}_i) \Longleftrightarrow$$

$$\Longleftrightarrow x \in \bigcup_{i \in Z} C(M_i) = \bigcup_{i \in Z} S_i^m.$$

Замечания. 1. Доказательство теоремы 2.9 может быть проведено и аналогично доказательству теоремы 2.8.

Избранный здесь путь сопоставления M и $\tilde{M}$ и переноса соответствующих утверждений о конусах на выпуклые множества называется "гомогенизацией".

2. Тот факт, что для множества $M \subseteq \mathbf{R}^n$, $\dim M = m$, действительно могут существовать точки $\bar{x} \in C(M)$, которые нельзя представить в виде выпуклой линейной комбинации меньшего чем $m + 1$ числа точек из M, иллюстрирует приведенный на рис. 2.9 пример для случая $\dim M = 2$.

Далее установим еще одно свойство симплекса.

Теорема 2.10. *Если V – линейное многообразие, S – симплекс, то $P = V \cap S$ – выпуклый многогранник.*

Доказательство. Пусть $S = C(\{a^i \in \mathbf{R}^n: i \in I = \{1, \ldots, r+1\}\})$, где векторы a^i аффинно независимы. В предположении $P \neq \phi$ каждой точке $v \in P$ поставим в соответствие симплекс наименьшей размерности вида

$$Sy = C(\{a^i \in \mathbf{R}^n: i \in J \subseteq I\}) \subseteq S$$

так, чтобы $y \in Sy$.

Всего имеется $2^{r+1} - 1$ симплексов Sy, вершины которых одновременно являются вершинами S. Если выполняется соотношение $Sy \cap P = \{y\}$, то y называется образующим вектором для P. Число образующих векторов конечно, поскольку каждое множество Sy имеет не более одного образующего вектора для P.

Покажем, что множество P является выпуклой оболочкой образующих векторов.

Если $y \in P$ не является образующим вектором, то существует $z \in \mathbf{R}^n$, $z \neq 0$, такой, что $y + z \in P \cap Sy$. Пусть

$$y = \sum_{i=1}^{s} \lambda_i a^i, \quad \lambda_i > 0, \quad i = 1, \ldots, s,$$

$$\sum_{i=1}^{s} \lambda_i = 1, \quad 1 \leqslant s \leqslant r + 1.$$

Тогда найдутся числа δ_i, $i = 1, \ldots, s$, $\sum_{i=1}^{s} \delta_i = 0$, такие, что выполняется равенство

$$y + z = \sum_{i=1}^{s} (\lambda_i + \delta_i) a^i.$$

Выберем теперь два числа μ_1 и μ_2 из соотношений

$$\mu_1 = \min \{-\lambda_i/\delta_i: \ \delta_i < 0\}, \quad \mu_2 = \min \{\lambda_i/\delta_i: \ \delta_i > 0\}$$

и образуем векторы

$$z^1 = y + \mu_1 z = \sum_{i=1}^{s} (\lambda_i + \mu_1 \delta_i) a^i, \tag{2.42а}$$

$$z^2 = y - \mu_2 z = \sum_{i=1}^{s} (\lambda_i - \mu_2 \delta_i) a^i. \tag{2.42б}$$

В силу определения чисел μ_1 и μ_2 в равенствах (2.42а) и (2.42б) по меньшей мере один из коэффициентов равен нулю. Кроме того,

$$\lambda_i + \mu_1 \delta_i \geqslant 0, \qquad i = 1, \ldots, s; \qquad \sum_{i=1}^{s} (\lambda_i + \mu_1 \delta_i) = 1,$$

$$\lambda_i - \mu_2 \delta_i \geqslant 0, \qquad i = 1, \ldots, s; \qquad \sum_{i=1}^{s} (\lambda_i - \mu_2 \delta_i) = 1.$$

Поэтому z^1 и z^2 принадлежат симплексам меньшей, чем у Sy, размерности, а y представим в виде выпуклой линейной комбинации z^1 и z^2:

$$y = \frac{1}{\mu_1 + \mu_2} (\mu_2 z^1 + \mu_1 z^2).$$

Если теперь z^1 и (или) z^2 все еще не являются образующими векторами, то поступаем с этими точками так же, как ранее с y. После конечного числа таких шагов y будет представлен в виде выпуклой линейной комбинации образующих векторов.

Из теоремы 2.10 следует, что пересечение D подпространства $L \subseteq \mathbf{R}^n$ и неотрицательного ортанта $\mathbf{R}^n_+$ может быть представлено в виде положительной оболочки конечного числа векторов. В самом деле, $D = L \cap \mathbf{R}^n_+ = K(L \cap P^n_+) = K^c(L \cap P^n_+)$, а $L \cap P^n_+$ по теореме 2.10 является выпуклым многогранником $P = L(a^1, \ldots, a^s)$.

Из линейной алгебры известно, что любое подпространство и любое линейное многообразие пространства $\mathbf{R}^n$ может быть представлено в виде линейной и соответственно аффинной оболочки конечного числа векторов. Аналогичные утверждения справедливы при применении понятий положи-

тельной оболочки и выпуклой оболочки к многогранным конусам и соответственно к многогранным множествам.

Т е о р е м а 2.11 (*теорема Минковского*). *Любой многогранный конус*

$$K = \{x \in \mathbf{R}^n:\ Ax \leqslant 0\},$$

где A *– матрица порядка* $m \times n$, *может быть представлен в виде*

$$K = \{x \in \mathbf{R}^n:\ x = Bz,\quad z \in \mathbf{R}^r_+\}, \tag{2.43}$$

т.е. K *является положительной оболочкой конечного числа векторов.*

Д о к а з а т е л ь с т в о. 1. Рассмотрим множества $U = \{(x, y)' \in \mathbf{R}^{n+m}.\ x \in \mathbf{R}^n, y \in \mathbf{R}^m_+\}$ и $V = \{(x, y)' \in \mathbf{R}^{n+m}:\ Ax + y = 0\}$. Заметим, что V – подпространство $\mathbf{R}^{n+m}$. Пусть g: $\mathbf{R}^{n+m} \to \mathbf{R}^n$ – линейное отображение, определенное соотношением $g\begin{pmatrix} x \\ y \end{pmatrix} = x$. Тогда, очевидно, $K = g(U \cap V)$.

Покажем, что существует конечное число векторов $\begin{pmatrix} b^i \\ y^i \end{pmatrix} \in \mathbf{R}^{n+m}$, $i = 1, \ldots, r$, таких, что

$$U \cap V = K^c\left(\left\{\begin{pmatrix} b^i \\ y^i \end{pmatrix}\right\}\right).$$

Отсюда сразу же будет следовать утверждение теоремы, поскольку

$$K = \{x:\ Ax \leqslant 0\} = g(U \cap V) = K^c\left(\left\{g\begin{pmatrix} b^i \\ y^i \end{pmatrix}\right\}\right) =$$

$$= K^c(\{b^i\}) = \{x:\ Bz = x,\ z \in \mathbf{R}^r_+\};\quad B = (b^1, \ldots, b^r).$$

2. Для доказательства существования указанного выше представления множества $U \cap V$ заметим, что

$$U = K^c(\{\{e^i\}, (-1, \ldots, -1, 0, \ldots, 0)'\}),$$

где e^i, $i = 1, \ldots, n + m$, – векторы *канонического базиса* пространства $\mathbf{R}^{n+m}$, а вектор $(-1, \ldots, -1, 0, \ldots, 0)'$ имеет ровно m последних нулевых координат.

Ради краткости записи обозначим входящие в это представление векторы через p^i, $i = 1, \ldots, n + m + 1$. Для произвольного отображения f: $\mathbf{R}^{n+m+1} \to \mathbf{R}^{n+m}$ справедливо равенство

$$U \cap V = f(f^-(U \cap V)) = f(f^-(U) \cap f^-(V)).$$

Пусть теперь f: $\mathbf{R}^{n+m+1} \to \mathbf{R}^{n+m}$ – то линейное отображение, которое каждому вектору $u = (u_1, \ldots, u_{n+m+1})' \in \mathbf{R}^{n+m+1}$ ставит в соответствие образ $f(u) = \sum\limits_{i=1}^{n+m+1} u_i p^i$. Для этого отображения $f^-(U) = \mathbf{R}^{n+m+1}_+$, а $f^-(V)$ является подпространством $\mathbf{R}^{n+m+1}$ (как прообраз подпространства).

По теореме 2.10 существует конечное число векторов $q^j \in \mathbf{R}^{n+m+1}$, $j = 1, \ldots, r$, для которых $f^-(U \cap V) = f^-(U) \cap f^-(V) = K^c(\{q^j\})$. Тогда

окончательно $U \cap V = K^c(\{f(q^j)\})$. Векторы $f(q^j)$ при этом в силу конструкции отображения f и свойств p^j имеют вид

$$f(q^j) = \begin{pmatrix} b^j \\ y^j \end{pmatrix}, \quad b^j \in \mathbf{R}^n, \quad y^j \in \mathbf{R}^m_+.$$

Замечание. Справедливо и обратное, а именно, любое множество вида (2.43) является многогранным конусом (см. ниже теорему 2.44).

Из теоремы Минковского, которая характеризует структуру множества решений однородной системы линейных неравенств, следует утверждение о множестве решений произвольной системы линейных неравенств.

Теорема 2.12. *Любое многогранное множество*

$$M = \{x \in \mathbf{R}^n: \ Ax \leqslant b\},$$

где A – матрица порядка $m \times n$, $b \in \mathbf{R}^m$, может быть представлено в виде

$$M = P + K,$$

где P – выпуклый многогранник, а K – положительная оболочка конечного числа векторов.

Доказательство. Соотношение $x \in M \subseteq \mathbf{R}^n$ имеет место тогда и только тогда, когда

$$\begin{pmatrix} x \\ 1 \end{pmatrix} \in \tilde{M} = \left\{ \begin{pmatrix} x \\ \xi \end{pmatrix} \in \mathbf{R}^{n+1}: \ (A, -b) \begin{pmatrix} x \\ \xi \end{pmatrix} \leqslant 0, \ \xi \geqslant 0 \right\}.$$

Множество $\tilde{M}$ является многогранным конусом, который по теореме Минковского представим в виде положительной оболочки конечного числа векторов $\tilde{a}^i$, $i \in I$. Без ограничения общности будем считать, что все векторы $\tilde{a}^i$, $i \in I$, у которых последняя координата отлична от нуля, имеют форму $\tilde{a}^i = \begin{pmatrix} a^i \\ 1 \end{pmatrix} \in \mathbf{R}^{n+1}$, так что

$$\tilde{M} = K^c\left(\left\{\begin{pmatrix} a^i \\ 1 \end{pmatrix} \in \mathbf{R}^{n+1}, \ i \in I_1\right\}\right) + K^c\left(\left\{\begin{pmatrix} a^i \\ 0 \end{pmatrix} \in \mathbf{R}^{n+1}, \ i \in I_2\right\}\right),$$

где $I_1 \cup I_2 = I$, $I_1 \cap I_2 = \phi$.

Отсюда имеем

$$x \in M \Leftrightarrow \begin{pmatrix} x \\ 1 \end{pmatrix} \in \tilde{M} \Leftrightarrow \left\{ \begin{array}{l} \begin{pmatrix} x \\ 1 \end{pmatrix} \in \sum\limits_{i \in I_1} \lambda_i \begin{pmatrix} a^i \\ 1 \end{pmatrix} + \sum\limits_{i \in I_2} \mu_i \begin{pmatrix} a^i \\ 0 \end{pmatrix} \\ \lambda_i \geqslant 0, \ i \in I_1; \ \mu_i \geqslant 0, \ i \in I_2 \end{array} \right\} \Leftrightarrow$$

$$\Leftrightarrow x \in C(\{a^i \in \mathbf{R}^n, \ i \in I_1\}) + K^c(\{a^i \in \mathbf{R}^n, \ i \in I_2\}).$$

Теоремы о представлении произвольного выпуклого множества имеют, естественно, более сложную структуру. Обсуждение этой проблемы можно найти в [74, 79].

Мы ограничимся следующим утверждением.

Теорема 2.13. *Пусть $C \subseteq \mathbf{R}^n$ – выпуклое, замкнутое, но не ограниченное множество. Существует по меньшей мере один луч $S = \{z: \ z = \lambda z^0,$*

$\lambda \geqslant 0\}$, *где* z^0 – *фиксированный вектор,* $z^0 \neq 0$, *для которого справедливо утверждение*

$$x \in C \Rightarrow x + S \subseteq C.$$

Доказательство. Так как C не ограничено, то существует последовательность $\{x^i\}$ такая, что $\|x^i\| \to \infty$, $x^i \neq 0$ и $x^i \in C$ для всех i. Последовательность z^i, $z^i := x^i/\|x^i\|$, содержится в компактном множестве $S_1 = \{z : \|z\| = 1\}$ и имеет поэтому хотя бы одну предельную точку z^0. Тогда существует сходящаяся к $z^0 \in S_1$ подпоследовательность. Без ограничения общности считаем, что $z^i \to z^0$.

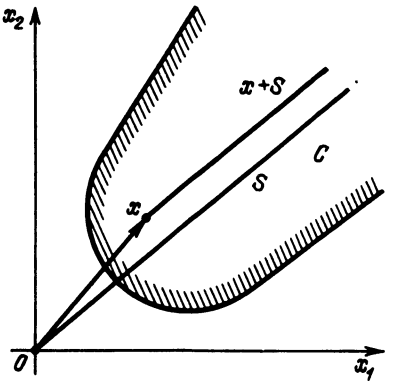


Рис. 2.10.

Для произвольного фиксированного $\lambda \geqslant 0$ и достаточно большого i выполняется неравенство

$$0 \leqslant \frac{\lambda}{\|x^i\|} \leqslant 1,$$

так что, с одной стороны, для достаточно большого i имеем $x + \frac{\lambda}{\|x^i\|}(x^i - x) \in C$, а с другой стороны, – последовательность $\left\{x + \frac{\lambda}{\|x^i\|}(x^i - x)\right\}$ сходится к $x + \lambda z^0$.

Теорема 2.13 проиллюстрирована рис. 2.10.

Важное комбинаторное свойство выпуклых множеств описывает следующая теорема.

Теорема 2.14 (*теорема Хелли*). *Пусть* $C_i \subseteq \mathbf{R}^n$, $i \in I$, $I = \{1, \ldots, m\}$, $m \geqslant n + 1$, – *выпуклые множества. Если каждые* $n + 1$ *множеств* C_i *имеют непустое пересечение, то и пересечение* $\bigcap_{i \in I} C_i$ *является непустым множеством.*

Доказательство. Проведем доказательство индукцией по m. При $m = n + 1$ утверждение тривиально.

Из предположения, что теорема верна при $m - 1 \geqslant n + 1$, следует существование m точек $x^i \in \bigcap_{\substack{k=1 \\ k \neq i}}^{m} C_k$, $i = 1, \ldots, m$. Тогда система $n + 1$ линейных уравнений с $m > n + 1$ неизвестными $\sum_{i=1}^{m} \lambda_i x^i = 0$, $\sum_{i=1}^{m} \lambda_i = 0$ имеет нетривиальное решение $(\lambda_1, \ldots, \lambda_m)' \neq 0$. Без ограничения общности можно считать, что $\lambda_1 \geqslant 0, \ldots, \lambda_r \geqslant 0$ и $\lambda_{r+1} \leqslant 0, \ldots, \lambda_m \leqslant 0$. Тогда

$$\sum_{i=1}^{r} \lambda_i = -\sum_{i=r+1}^{m} \lambda_i > 0.$$

Так как $x^i \in \bigcap_{k=r+1}^{m} C_k$ для каждого $i \in \{1, \ldots, r\}$, то выпуклая линейная комбинация $x = \left(\sum_{i=1}^{r} \lambda_i x^i\right)\left(\sum_{i=1}^{r} \lambda_i\right)^{-1}$ точек x^i также принадлежит множеству $\bigcap_{k=r+1}^{m} C_k$. Но тогда в силу предположения индукции имеем

$$\frac{\sum_{i=r+1}^{m} \lambda_i x^i}{\sum_{i=r+1}^{m} \lambda_i} = \frac{-\sum_{i=1}^{r} \lambda_i x^i}{\left(-\sum_{i=1}^{r} \lambda_i\right)} = x \in \bigcap_{k=1}^{r} C_k,$$

и окончательно $x \in \bigcap_{k=1}^{m} C_k$, т.е. $\bigcap_{k=1}^{m} C_k \neq \phi$.

В заключение приведем одно утверждение типа теоремы Хелли, относящееся к компактным выпуклым множествам и означающее тем самым переход к следующему пункту, в котором рассматриваются топологические свойства выпуклых множеств.

Т е о р е м а 2.15. *Пусть* $(C_i)_{i \in Z}$ – *произвольное семейство компактных выпуклых множеств* $C_i \subseteq \mathbf{R}^n$. *Если каждые* $n + 1$ *множеств* C_i *имеют непустое пересечение, то и пересечение* $\bigcap_{i \in Z} C_i$ *не пусто.*

Д о к а з а т е л ь с т в о. Предположим, что $\bigcap_{i \in Z} C_i = \phi$. Тогда

$$\bigcup_{i \in Z} \tilde{C}_i = \mathbf{R}^n,$$

где $\tilde{C}_i = \mathbf{R}^n \setminus C_i$ – дополнение к множеству C_i. В частности, для компактного множества C_{i_*}, $i_* \in Z$, выполняется

$$\bigcup_{i \in Z} \tilde{C}_i \supseteq C_{i_*}.$$

Открытые множества $\tilde{C}_i$, $i \in Z$, образуют покрытие C_{i_*}. Так как C_{i_*} компактно, то по теореме Гейне – Бореля существует конечное покрытие множества C_{i_*}, состоящее из множеств $\tilde{C}_{ij}$, $i \in Z$:

$$\bigcup_{j=1}^{r} \tilde{C}_{ij} \supseteq C_{i_*}.$$

Отсюда вытекает, что $\bigcap_{j=1}^{r} C_{ij} \subseteq \tilde{C}_{i_*}$ и $\bigcap_{j=1}^{r} C_{ij} \cap C_{i_*} = \varphi$. Последнее, однако, противоречит теореме 2.14, согласно которой любое конечное подсемейство семейства $(C_i)_{i \in Z}$ имеет непустое пересечение.

Таким образом, $\bigcap_{i \in Z} C_i \neq \phi$.

2.1.3. Топологические свойства выпуклых множеств. Наряду с введенными в п. 2.1.2 операциями взятия оболочки рассмотрим далее аналогичную операцию взятия замкнутой оболочки множества.

Замкнутой оболочкой $\bar{M}$ множества $M \subseteq \mathbf{R}^n$ называется пересечение всех замкнутых множеств, содержащих M. Как известно, $\bar{M}$ совпадает с множеством точек прикосновения M.

Используя единичный шар $S_1 = S(0, 1)$, получим следующее представление для $\bar{M}$:

$$\bar{M} = \bigcap_{\epsilon > 0} (M + \epsilon S_1). \tag{2.44}$$

В самом деле, элемент $x \in \mathbf{R}^n$ тогда и только тогда является точкой прикосновения множества M, когда $(x + \epsilon S_1) \cap M \neq \phi \quad \forall \epsilon > 0$. Таким образом,

$$\bar{M} = \{x: \ (x + \epsilon S_1) \cap M \neq \phi \quad \forall \epsilon > 0\}.$$

В силу теоремы 2.2 и свойства $S_1 = -S_1$ следует

$$\bar{M} = \{x: \{x\} \cap (M + \epsilon S_1) \neq \phi \quad \forall \epsilon > 0\} = \bigcap_{\epsilon > 0} (M + \epsilon S_1).$$

Внутренностью множества M называется множество

$$\operatorname{int} M = \{x \in M: \ \exists \epsilon > 0, \ \ x + \epsilon S_1 \subseteq M\}. \tag{2.45a}$$

Далее мы приведем некоторые свойства множеств, в которых не используется понятие выпуклости. Эти свойства потребуются, однако, для последующих рассуждений. Так, при доказательстве теорем отделимости для выпуклых множеств в п. 2.1.4 мы используем определенные топологические свойства суммы двух множеств.

Т е о р е м а 2.16. 1. *Пусть* $M_1, M_2 \subseteq \mathbf{R}^n$ *и множество* M_2 *открыто. Тогда открыто и множество* $M_1 + M_2$.

2. *Пусть* $M_1, M_2 \subseteq \mathbf{R}^n$ *и множество* M_1 *замкнуто, а* M_2 *компактно. Тогда множество* $M_1 + M_2$ *замкнуто.*

Следующий пример показывает, что сумма двух замкнутых множеств не всегда замкнута.

П р и м е р 2.7. Для замкнутых множеств

$$M_1 = \{x = (x_1, x_2)' \in \mathbf{R}^2: \ x_2 \geqslant e^{x_1}\},$$

$$M_2 = \{x = (x_1, x_2)' \in \mathbf{R}^2: \ x_2 = 0, \ x_1 \geqslant 0\}$$

имеет место равенство (рис. 2.11)

$$M_1 + M_2 = \{x \in \mathbf{R}^2: \ x_2 > 0\}.$$

Вообще говоря, не замкнута даже сумма двух замкнутых выпуклых конусов.

П р и м е р 2.8. Для замкнутых выпуклых конусов

$$K_1 = \{x = (x_1, x_2, x_3)' \in \mathbf{R}^3: \ x_1 + x_3 \geqslant \sqrt{x_1^2 + x_2^2 + x_3^2}\},$$

$$K_2 = \{x = (x_1, x_2, x_3)' \in \mathbf{R}^3: \ x_1 = 0, \ x_2 \leqslant 0, \ x_3 = 0\}$$

имеет место равенство (рис. 2.12)

$$K_1 + K_2 = \{x \in \mathbf{R}^3: \ x_3 > 0\} \cup \{x \in \mathbf{R}^3: \ x_1 = 0, \ x_3 \geqslant 0\}.$$

Напротив, сумма многогранных конусов всегда замкнута. Это следует из теоремы 2.11, если учесть, что положительная оболочка конечного числа векторов является замкнутым множеством. Последнее утверждение составляет содержание следующей теоремы.

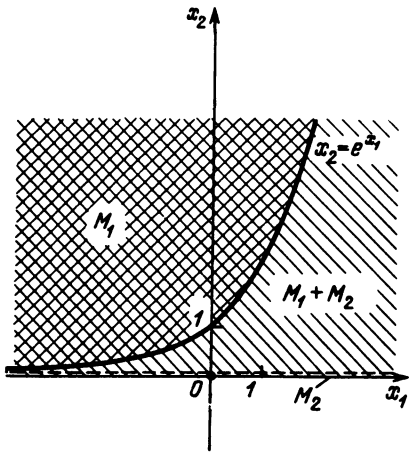


Рис. 2.11.

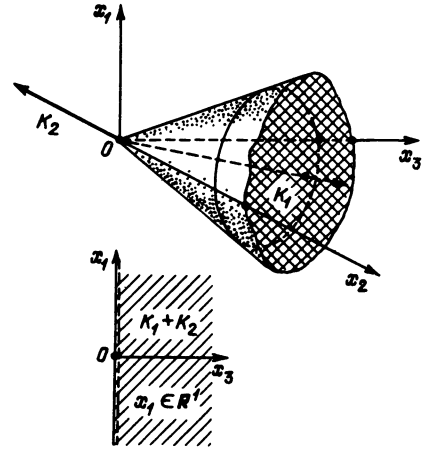


Рис. 2.12.

Т е о р е м а 2.17. *Пусть* $B = (b^1, \ldots, b^r)$ – *матрица порядка* $n \times r$. *Тогда множество*

$$K = \{x \in \mathbf{R}^n: \ x = Bz, \ z \in \mathbf{R}^r_+\}$$

замкнуто.

Д о к а з а т е л ь с т в о. 1. Множество K является положительной оболочкой $M = \{b^1, \ldots, b^r\}$. Из M можно выбрать лишь конечное число отличных друг от друга подмножеств M_s, каждое из которых состоит из s $(1 \leqslant s \leqslant r)$ линейно независимых векторов. Покажем, что положительная оболочка $K^c(M_s)$ s линейно независимых векторов есть замкнутое множество. Без ограничения общности можно считать, что $b^1, \ldots, b^s$ $(1 \leqslant s \leqslant r)$ линейно независимы. Тогда найдутся векторы $a^{s+1}, \ldots, a^n$ такие, что векторы-столбцы матрицы $B = (b^1, \ldots, b^s, a^{s+1}, \ldots, a^n)$ образуют базис пространства $\mathbf{R}^n$. Имеем

$$K^c(M_s) = \{x: \ x = Dz; \ z \geqslant 0, \ i = 1, \ldots, s; \ z_i = 0, \ i = s+1, \ldots, n\}.$$

В силу непрерывности линейного отображения $z = D^{-1}x$ прообраз $K^c(M_s)$ замкнутого множества

$$\{z \in \mathbf{R}^n: \ z_i \geqslant 0, \ i = 1, \ldots, s; \ z_i = 0, \ i = s+1, \ldots, n\}$$

замкнут.

2. Объединение множеств $K^c(M_s)$ по $s = 1, \ldots, r$ дает согласно теореме 2.8 множество K, которое (как объединение конечного числа замкнутых множеств) замкнуто.

При линейном преобразовании A: $\mathbf{R}^n \to \mathbf{R}^m$ замкнутое множество не обязательно отображается на замкнутое множество.

Т е о р е м а 2.18. *Пусть* $M \subseteq \mathbf{R}^n$ – *произвольное множество и* A: $\mathbf{R}^n \to \mathbf{R}^m$ – *линейное преобразование. Тогда*

$$\overline{AM} \subseteq \overline{AM}.$$

Доказательство. С одной стороны,

$$\overline{AM} = \{y: \exists\{x^i\} \subseteq M,\ Ax^i \to y\}.$$

С другой, в силу непрерывности линейного преобразования, имеем

$$A\bar{M} = \{z: \exists\{x^i\} \subseteq M,\ x^i \to x,\ z = Ax\}.$$

Для каждого $z \in A\bar{M}$ ввиду $Ax^i \to Ax$ при $x^i \to x$ выполняется также $z \in \overline{AM}$.

Следующий пример иллюстрирует тот факт, что полученный с помощью линейного преобразования A образ AM замкнутого множества M действительно не обязан быть замкнутым.

Пример 2.9. Рассмотрим замкнутое выпуклое множество

$$C = \{x = (x_1, x_2, x_3)' \in \mathbf{R}^3: x_1 x_3 \geqslant 1,\ x_1 > 0\}$$

и матрицу проектирования

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Проекцией AC множества C на подпространство $L = \{x \in \mathbf{R}^3: x_3 = 0\}$ является множество $AC = \{x \in \mathbf{R}^3: x_1 > 0,\ x_3 = 0\}$. Очевидно, что множество AC не замкнуто.

Компактное множество из $\mathbf{R}^n$ линейное преобразование всегда переводит снова в компактное множество. Это вытекает из следующей известной теоремы.

Теорема 2.19. *Пусть $M \subseteq \mathbf{R}^n$ – компактное множество и f: $\mathbf{R}^n \to \mathbf{R}^m$ – непрерывное отображение. Тогда множество $f(M)$ компактно.*

С помощью этого утверждения докажем следующую теорему.

Теорема 2.20. *Выпуклая оболочка конечного числа векторов есть компактное множество.*

Доказательство. Пусть $M = \{a^1, a^2, \ldots, a^m\}$. Тогда

$$C(M) = \{x: x = \sum_{i=1}^{m} \lambda_i a^i,\ (\lambda_1, \ldots, \lambda_m)' \in P_+^m\}.$$

Множество P_+^m замкнуто как пересечение замкнутых множеств, и поскольку $\lambda \in P_+^m$ влечет $\|\lambda\| \leqslant 1$, где $\lambda = (\lambda_1, \ldots, \lambda_m)'$, оно также и компактно.

По теореме 2.19 множество $C(M)$ компактно как образ компактного множества P_+^m при задаваемом матрицей $A = (a^1, \ldots, a^m)$ линейном отображении.

Обсудим теперь некоторые топологические свойства выпуклых множеств.

Теорема 2.21. *Если $C \subseteq \mathbf{R}^n$ – выпуклое множество, то и множества $\bar{C}$ и $\operatorname{int} C$ выпуклые.*

Доказательство. Используем представление $\bar{C}$ и $\operatorname{int} C$ согласно (2.44) и (2.45а) при $M = C$.

1. Очевидно, что для произвольного $\epsilon > 0$ множество $C + \epsilon S_1$ выпукло, следовательно, по теореме 2.3 выпукло и множество $\bigcap_{\epsilon > 0} (C + \epsilon S_1)$.

2. Справедливы соотношения

$$x, y \in \operatorname{int} C \Rightarrow \left\{ \begin{array}{ll} \exists \epsilon_1 > 0: & x + \epsilon_1 S_1 \subseteq C \\ \exists \epsilon_2 > 0: & y + \epsilon_2 S_1 \subseteq C \end{array} \right\} \Rightarrow$$

$$\Rightarrow \lambda_1 x + \lambda_2 y + (\lambda_1 \epsilon_1 + \lambda_2 \epsilon_2) S_1 \subseteq C,$$

где $(\lambda_1, \lambda_2)' \in P_+^2$. Таким образом, для точки $z = \lambda_1 x + \lambda_2 y$, где $(\lambda_1, \lambda_2)' \in P_+^2$, существует $\epsilon = \lambda_1 \epsilon_1 + \lambda_2 \epsilon_2 > 0$ такое, что $z + \epsilon S_1 \subseteq C$ и, тем самым, $z \in \operatorname{int} C$.

Часто нет необходимости соотносить внутренность множества M со всем пространством $\mathbf{R}^n$. Вместо этого рассматривают внутренность относительно аффинной оболочки $V(M)$ множества M и говорят об относительной внутренности M (относительно $V(M)$).

Точка $x^0 \in M$ называется *относительно внутренней точкой* M, если x^0 – внутренняя точка относительно топологии в аффинной оболочке $V(M)$ множества M, индуцированной топологией в $\mathbf{R}^n$ *).

Множество

$$\operatorname{ri} M = \{x \in M: \exists \epsilon > 0, \ (x + \epsilon S_1) \cap V(M) \subseteq M\} \qquad (2.45б)$$

всех относительно внутренних точек M назовем *относительной внутренностью множества* M **).

Очевидно, что $\operatorname{ri} M \subseteq M \subseteq \bar{M}$.

Для n-мерного множества $M \subseteq \mathbf{R}^n$ справедливо, в частности, $V(M) = \mathbf{R}^n$, и поэтому $\operatorname{ri} M = \operatorname{int} M$.

Множество M называется *относительно открытым* (относительно индуцированной топологии в $V(M)$), если $\operatorname{ri} M = M$.

Множество $\bar{M} \setminus \operatorname{ri} M$ называется *относительной границей* M, любая точка из $\bar{M} \setminus \operatorname{ri} M$ – *относительно граничной точкой* M ***).

При доказательстве утверждений, касающихся относительной внутренности множества $M \subseteq \mathbf{R}^n$, без ограничения общности можно считать, что M имеет внутренние точки. Дело в том, что в случае $\dim M = m < n$ с помощью взаимно однозначного аффинного преобразования можно множество M отобразить на подпространство, определяемое равенствами $x_{m+1} = 0, \ldots$ $\ldots, x_n = 0$, которое изоморфно и изометрично $\mathbf{R}^m$.

Т е о р е м а 2.22. *Если* $C \subseteq \mathbf{R}^n$ – *непустое выпуклое множество, то* $\operatorname{ri} C \neq \phi$.

Д о к а з а т е л ь с т в о. Если $\dim C = 0$, то $C = \{x^0\}$ и, как мы условились, x^0 – относительно внутренняя точка C.

В случае $\dim C = m > 0$ без ограничения общности будем считать $m = n$.

С помощью аффинного преобразования всегда можно достичь того, чтобы содержащийся в C n-мерный симплекс S_n имел вершинами точки

---

*) Открытый шар с центром $x^0 \in V(M)$ и радиусом $\alpha > 0$ (шаровая окрестность точки x^0) в этой топологии определяется как $S^{V(M)}(x^0, \alpha) = V(M) \cap S(x^0, \alpha)$.

**) В случае $M = \{x^0\}$ точка x^0 рассматривается как относительно внутренняя точка множества M.

***) Это определение отличается от обычного определения граничной точки, однако оно оказывается целесообразным.

$(0, 0, \ldots, 0)$, $(1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 1)$. Тогда

$$S_n = \{x = (x_1, \ldots, x_n)' : \ x_1, \ldots, x_n \geqslant 0, \quad \sum_{i=1}^{n} x_i \leqslant 1\},$$

$$\operatorname{int} S_n = \{x = (x_1, \ldots, x_n)' : \ x_1, \ldots, x_n > 0, \quad \sum_{i=1}^{n} x_i < 1\} \neq \phi,$$

а отсюда следует $\operatorname{int} C \neq \phi$.

Для невыпуклых множеств утверждение теоремы 2.22, вообще говоря, неверно. Достаточно вспомнить хотя бы об окружности в $\mathbf{R}^2$.

Теорема 2.23. *Если* $C \subseteq \mathbf{R}^n$ *– выпуклое множество, то из соотношений* $x \in \operatorname{ri} C$, $y \in \bar{C}$ *следует* $\lambda_1 x + \lambda_2 y \in \operatorname{ri} C$, *где* $(\lambda_1, \lambda_2)' \in \operatorname{ri} P_+^2$.

Доказательство. Без потери общности проведем доказательство для $\dim C = n$, так что будем считать, что $\operatorname{ri} C = \operatorname{int} C$.

В силу $y \in \bar{C} \Rightarrow y \in C + \epsilon S_1 \quad \forall \epsilon > 0$ имеем

$$\lambda_1 x + \lambda_2 y + \epsilon S_1 \subseteq \lambda_1 [x + \epsilon(1 + \lambda_2)\lambda_1^{-1} S_1] + \lambda_2 C \quad \forall \epsilon > 0.$$

Полагая $\epsilon'(\epsilon) = \epsilon(1 + \lambda_2)\lambda_1^{-1}$ и учитывая $x \in \operatorname{ri} C$, получим $x + \epsilon'(\epsilon) S_1 \subseteq C$ $\forall \epsilon \in [0, \epsilon_0)$, и поэтому $\lambda_1 x + \lambda_2 y + \epsilon S_1 \subseteq \lambda_1 C + \lambda_2 C = C$. Отсюда для $\epsilon \in \in [0, \epsilon_0)$ следует

$$\lambda_1 x + \lambda_2 y + \epsilon S_1 \subseteq C \quad \forall (\lambda_1, \lambda_2)' \in \operatorname{ri} P_+^2.$$

В качестве непосредственного следствия из предыдущего утверждения получим следующее усиление теоремы 2.21.

Теорема 2.24. *Если* $C \subseteq \mathbf{R}^n$ *выпукло, то выпукло и* $\operatorname{ri} C$.

Доказательство. Вместо $y \in \bar{C}$ в теореме 2.23 возьмем $y \in \operatorname{ri} C$.

Теорема 2.25. *Пусть* $C \subseteq \mathbf{R}^n$ *– выпуклое множество. Тогда*

$\overline{\operatorname{ri} C} = \bar{C}$, $\operatorname{ri} \bar{C} = \operatorname{ri} C$.

Доказательство. 1. С одной стороны, в силу $\operatorname{ri} C \subseteq C$ имеем $\overline{\operatorname{ri} C} \subseteq \bar{C}$. С другой стороны, в случае $C \neq \phi$ для заданного $y \in \bar{C}$ и $x \in \operatorname{ri} C$ согласно теореме 2.23 получим

$$\lambda x + (1 - \lambda) y \in \operatorname{ri} C \quad \forall \lambda \in (0, 1].$$

Поэтому $y \in \overline{\operatorname{ri} C}$ и, следовательно, $\bar{C} \subseteq \overline{\operatorname{ri} C}$.

Отсюда получаем $\overline{\operatorname{ri} C} = \bar{C}$.

2. В силу $C \subseteq \bar{C}$ и $V(\bar{C}) = V(C)$ имеем $\operatorname{ri} C \subseteq \operatorname{ri} \bar{C}$. Пусть теперь $\bar{x} \in \operatorname{ri} \bar{C}$, x – некоторая точка из $\operatorname{ri} C$, $x \neq \bar{x}$.

Элементы x и $\bar{x}$ определяют прямую

$$\{y : y = \lambda \bar{x} + (1 - \lambda) x = \bar{x} + (1 - \lambda)(x - \bar{x}); \lambda \in \mathbf{R}\},$$

точки которой при $\lambda > 1$ и достаточно малом $\lambda - 1$ принадлежат множеству $\operatorname{ri} \bar{C}$, а, следовательно, и множеству $\bar{C}$. Для такого y точка $\bar{x}$ может быть записана в виде $\bar{x} = \mu y + (1 - \mu) x$, $\mu = \lambda^{-1} \in (0, 1]$. Тогда из теоремы 2.23 следует, что $\bar{x} \in \operatorname{ri} C$. Таким образом, мы получили $\operatorname{ri} \bar{C} \subseteq \operatorname{ri} C$, и поэтому $\operatorname{ri} \bar{C} = \operatorname{ri} C$.

Как непосредственное следствие из теоремы 2.23 вытекает следующий результат.

Теорема 2.26. *Пусть $C_1, C_2 \subseteq \mathbf{R}^n$ – выпуклые множества; тогда следующие соотношения эквивалентны:*

1) $\overline{C}_1 = \overline{C}_2$;

2) $\mathrm{ri}\, C_1 = \mathrm{ri}\, C_2$;

3) $\mathrm{ri}\, C_1 \subseteq C_2 \subseteq \overline{C}_1$.

Доказательство. 1. В силу теоремы 2.25 из равенства $\overline{C}_1 = \overline{C}_2$ следует

$$\mathrm{ri}\, C_1 = \mathrm{ri}\,\overline{C}_1 = \mathrm{ri}\,\overline{C}_2 = \mathrm{ri}\, C_2.$$

2. Из того, что $\mathrm{ri}\, C_1 = \mathrm{ri}\, C_2$, вытекает

$$\mathrm{ri}\, C_1 = \mathrm{ri}\, C_2 \subseteq C_2 \subseteq \overline{C}_2 = \overline{\mathrm{ri}\, C_2} = \overline{\mathrm{ri}\, C_1} = \overline{C}_1.$$

3. Наконец, из соотношения $\mathrm{ri}\, C_1 \subseteq C_2 \subseteq \overline{C}_1$ следует $\overline{C}_1 = \overline{\mathrm{ri}\, C_1} \subseteq \overline{C}_2 \subseteq \overline{C}_1$, и отсюда $\overline{C}_1 = \overline{C}_2$.

Замечание. Из $C_1 \subseteq C_2$ следует $\overline{C}_1 \subseteq \overline{C}_2$ и $\mathrm{int}\, C_1 \subseteq \mathrm{int}\, C_2$, но не $\mathrm{ri}\, C_1 \subseteq \mathrm{ri}\, C_2$, в чем можно непосредственно убедиться на простых примерах.

Теорема 2.27. *Пусть $C \subseteq \mathbf{R}^n$ – непустое выпуклое множество. Справедливы утверждения:*

1) $x \in \mathrm{ri}\, C \Leftrightarrow \forall y \in C \;\; \exists \mu > 1: \;\; \mu x + (1-\mu) y \in C$;

2) $x \in \mathrm{int}\, C \Leftrightarrow \forall y \in \mathbf{R}^n \;\; \exists \epsilon > 0: \;\; x + \epsilon y \in C$.

Доказательство. 1. Утверждение ($\Rightarrow$) следует непосредственно из определения (2.45б) относительной внутренности, поэтому мы ограничимся доказательством справедливости импликации ($\Leftarrow$). Так как $\mathrm{ri}\, C \neq \phi$ (теорема 2.22), то существует $\bar{y} \in \mathrm{ri}\, C$. Пусть $\bar{y} \neq x$ (если $\bar{y} = x$, то требуемое доказано). По предположению для $\bar{y} \in \mathrm{ri}\, C$ существует точка $z \in C$ такая, что $z = \mu x + (1-\mu)\bar{y}$ для некоторого $\mu > 1$. Тогда $x = \lambda z + (1-\lambda)\bar{y}$, где $\lambda = \mu^{-1} \in (0,1), z \in C, \bar{y} \in \mathrm{ri}\, C$, и поэтому согласно теореме 2.23 $x \in \mathrm{ri}\, C$.

2. Это утверждение непосредственно следует из доказанного выше.

Теорема 2.28. *Пусть $(C_i)_{i \in I}$ – семейство выпуклых множеств $C_i \subseteq \mathbf{R}^n$ такое, что $\bigcap\limits_{i \in I} (\mathrm{ri}\, C_i) \neq \phi$. Тогда*

$$\overline{\bigcap_{i \in I} C_i} = \bigcap_{i \in I} \overline{C}_i. \tag{2.46}$$

В случае конечного множества индексов $I = \{1, \ldots, k\}$ имеет место соотношение

$$\mathrm{ri}\Big(\bigcap_{i=1}^{k} C_i\Big) = \bigcap_{i=1}^{k} (\mathrm{ri}\, C_i). \tag{2.47}$$

Доказательство. 1. Пусть $x \in \bigcap\limits_{i \in I} (\mathrm{ri}\, C_i)$ и $y \in \bigcap\limits_{i \in I} \overline{C}_i$. Тогда по теореме 2.23 имеем

$$\lambda x + (1-\lambda) y \in \bigcap_{i \in I} (\mathrm{ri}\, C_i) \;\; \forall \lambda \in (0,1].$$

При $\lambda \to +0$ элемент y оказывается предельной точкой векторов $\lambda x + (1-\lambda) y$, так что

$$\bigcap_{i \in I} \overline{C}_i \subseteq \overline{\bigcap_{i \in I} (\mathrm{ri}\, C_i)} \subseteq \overline{\bigcap_{i \in I} C_i} \subseteq \bigcap_{i \in I} \overline{C}_i.$$

Этим и доказано (2.46).

2. Одновременно с (2.46) выше было показано, что $\overline{\bigcap_{i\in I} (\mathrm{ri}\, C_i)} = \overline{\bigcap_{i\in I} C_i}$. По теореме 2.26 тогда выполняется также равенство

$$\mathrm{ri}(\bigcap_{i\in I} \mathrm{ri} C_i) = \mathrm{ri}(\bigcap_{i\in I} C_i),$$

и поэтому

$$\mathrm{ri}(\bigcap_{i\in I} C_i) \subseteq \bigcap_{i\in I} \mathrm{ri} C_i. \tag{2.48}$$

Пусть теперь $I = \{1, \ldots, k\}$ и $y \in \bigcap_{i=1}^{k} \mathrm{ri} C_i$. В множестве $\bigcap_{i=1}^{k} C_i$ любой отрезок прямой с граничной точкой y может быть продолжен из y в каждое из множеств C_i. Поскольку I конечно, пересечение таких продолжений будет продолжением первоначального сегмента в множестве $\bigcap_{i=1}^{k} C_i$. Тогда по теореме 2.27 имеем $y \in \mathrm{ri}(\bigcap_{i=1}^{k} C_i)$, и окончательно $\bigcap_{i=1}^{k} \mathrm{ri} C_i \subseteq \mathrm{ri}(\bigcap_{i=1}^{k} C_i)$. Отсюда и из (2.48) вытекает (2.47).

Теорема 2.29. *Пусть $C \subseteq \mathbf{R}^n$ – выпуклое множество и $A: \mathbf{R}^n \to \mathbf{R}^m$ – линейное преобразование. Тогда*

$$\mathrm{ri}(AC) = A(\mathrm{ri} C). \tag{2.49}$$

Доказательство. В силу теорем 2.25 и 2.18 имеем

$$\mathrm{ri}(A\,\mathrm{ri} C) \subseteq A(\mathrm{ri} C) \subseteq AC \subseteq A\overline{C} = A(\overline{\mathrm{ri} C}) \subseteq \overline{A(\mathrm{ri} C)}.$$

Отсюда согласно теореме 2.26 следует $\overline{AC} = \overline{A(\mathrm{ri} C)}$, а потому $\mathrm{ri}(AC) = \mathrm{ri}(A(\mathrm{ri} C))$ и, наконец,

$$\mathrm{ri}(AC) \subseteq A(\mathrm{ri} C). \tag{2.50}$$

Пусть теперь $z \in A(\mathrm{ri} C)$, $x \in AC$, а $z' \in \mathrm{ri} C$, $x' \in C$ – некоторые прообразы z и x соответственно, так что $Az' = z$, $Ax' = x$. Тогда по теореме 2.27 существует число $\mu > 1$ такое, что $y = \mu z' + (1 - \mu)x' \in C$. Для образа точки y при отображении A при том же $\mu > 1$ имеем $\mu z + (1-\mu)\, x \in AC$, и поэтому $z \in \mathrm{ri}(AC)$. Таким образом, мы получаем $A(\mathrm{ri} C) \subseteq \mathrm{ri}(AC)$, что вместе с (2.50) и дает (2.49).

Теорема 2.30. *Пусть $C_1 \subseteq \mathbf{R}^{n_1}$ и $C_2 \subseteq \mathbf{R}^{n_2}$ – выпуклые множества. Тогда*

$$\overline{C_1 \times C_2} = \overline{C}_1 \times \overline{C}_2, \quad \mathrm{ri}(C_1 \times C_2) = \mathrm{ri} C_1 \times \mathrm{ri} C_2.$$

Доказательство следует из определения декартова произведения двух множеств.

Теорема 2.31. *Пусть $C_1, C_2 \subseteq \mathbf{R}^n$ – выпуклые множества. Тогда*

$$\overline{C_1 + C_2} \supseteq \overline{C}_1 + \overline{C}_2, \quad \mathrm{ri}(C_1 + C_2) = \mathrm{ri} C_1 + \mathrm{ri} C_2.$$

Доказательство. Применим линейное преобразование $A: \mathbf{R}^{2n} \to \mathbf{R}^n$ вида $A(x^1, x^2) = x^1 + x^2 \;\; \forall x^1, x^2 \in \mathbf{R}^n$. Тогда $C_1 + C_2 = A(C_1 \times C_2)$. Требуемое утверждение следует из теорем 2.18 и 2.29.

2.1.4. Теоремы отделимости. Если даны два непересекающихся выпуклых множества C_1 и C_2, то кажется очевидным, что они могут быть "отделены" друг от друга некоторой гиперплоскостью. Итак, возникает предположение о существовании такой гиперплоскости, что множество C_1 полностью лежит по одну сторону, а C_2 – полностью по другую сторону от этой гиперплоскости. Утверждения подобного рода объединяются под названием теорем отделимости и играют важную роль в теории выпуклых множеств и в нелинейном программировании.

Предварительно для произвольных множеств введем понятия *разделяющей гиперплоскости* и *строго разделяющей гиперплоскости.* Гиперплоскость $E(a, \alpha) = \{x \in \mathbf{R}^n: a'x = \alpha\}$ разделяет два непустых множества M_1, $M_2 \subseteq \mathbf{R}^n$, если

$$a'x \leqslant \alpha \quad \forall x \in M_1, \qquad a'x \geqslant \alpha \quad \forall x \in M_2.$$

Гиперплоскость $E(a, \alpha)$ строго разделяет два непустых множества M_1, $M_2 \subseteq \mathbf{R}^n$, если

$$a'x \leqslant \alpha \quad \forall x \in M_1, \qquad a'x > \alpha \quad \forall x \in M_2.$$

Если разделяющая гиперплоскость $E(a, \alpha)$ не содержит одновременно оба разделяемые ею выпуклые множества C_1 и C_2, то $E(a, \alpha)$ называется *собственно разделяющей гиперплоскостью* для C_1 и C_2. Следовательно, в этом случае $E(a, \alpha) \not\supseteq C_1 \cup C_2$. Очевидно, строгая разделимость влечет собственную.

Пример 2.10. Два шара

$$S_1 = \bar{S}\left(\begin{pmatrix}2\\0\end{pmatrix}, 1\right), \quad S_2 = \bar{S}\left(\begin{pmatrix}0\\0\end{pmatrix}, 1\right)$$

собственно (но не строго) разделяются прямой $E(\alpha, a)$, где

$a = (1, 0)'$, $\alpha = 1$.

Вначале мы докажем две вспомогательные теоремы о разделяющих гиперплоскостях.

Теорема 2.32. *Пусть $C \subseteq \mathbf{R}^n$ – непустое замкнутое выпуклое множество и $0 \notin C$. Тогда существует гиперплоскость $E(a, \alpha)$ с $\alpha > 0$ такая, что $a'x > \alpha \quad \forall x \in C$.*

Доказательство. Пусть $\bar{S}$ – замкнутый шар с центром в нуле, который пересекает C. Так как множество $C \cap \bar{S}$ компактно, то непрерывная функция $g(x) = \|x\|$, $x \in C \cap \bar{S}$, достигает на этом множестве минимума в некоторой точке x^0. Тогда $\|x\| \geqslant \|x^0\| \quad \forall x \in C \cap \bar{S}$, и поэтому также $\|x\| \geqslant \|x^0\| \quad \forall x \in C$.

Поскольку $0 \notin C$, то $\|x^0\| > 0$. Положим $a = x^0$, $\alpha = \|x^0\|^2/2$ и покажем, что $x'x^0 \geqslant \|x^0\|^2 > \alpha \quad \forall x \in C$.

Ввиду $\|x\| \geqslant \|x^0\| \quad \forall x \in C$ и выпуклости C выполняется также неравенство $\|x^0 + \lambda[x - x^0]\| \geqslant \|x^0\| \quad \forall x \in C, \lambda \in (0, 1]$, а поэтому

$$(x^0 + \lambda[x - x^0])'(x^0 + \lambda[x - x^0]) = \|x^0 + \lambda[x - x^0]\|^2 \geqslant \|x^0\|^2 = x^{0\prime}x^0$$

и, далее,

$$2(x - x^0)'x^0 + \lambda(x - x^0)'(x - x^0) \geqslant 0. \tag{2.51}$$

Отсюда при $\lambda \to +0$ следует $(x - x^0)'x^0 \geqslant 0 \quad \forall x \in C$. Поэтому $x^{0\prime}x \geqslant x^{0\prime}x^0 = \|x^0\|^2 > \alpha$.

Теорема 2.33. *Пусть* $C \subseteq \mathbf{R}^n$ *– непустое выпуклое множество и* $0 \notin \operatorname{ri} C$. *Тогда существует гиперплоскость* $E(a,0)$ *такая, что* $a'x \geqslant 0 \ \forall x \in \bar{C}$ *и* $a'x > 0$ *для хотя бы одного* $x \in C$.

Доказательство. 1. Пусть M– произвольное конечное подмножество из $\operatorname{ri} C$. По теореме 2.20 множество $C(M)$ компактно. Кроме того, очевидно, $0 \notin C(M)$. Поэтому в силу теоремы 2.32 существует точка $y \in A = \{y: \|y\| = 1\}$ такая, что $y'x \geqslant 0 \ \forall x \in C(M)$. Отсюда для компактных множеств $A_x = \{y: \|y\| = 1, \ y'x \geqslant 0\}$ следует $\bigcap\limits_{x \in C} A_x \neq \phi$.

В самом деле, если предположить $\bigcap\limits_{x \in C} A_x = \phi$, то $\bigcup\limits_{x \in C} \tilde{A}_x \supseteq A$, т.е. множества $\tilde{A}_x, x \in C$, образуют открытое покрытие компактного множества A. По теореме Гейне – Бореля можно выделить конечное подпокрытие множества A; таким образом, существует конечное множество $M \subseteq C$, для которого $\bigcup\limits_{x \in M} \tilde{A}_x \supseteq A$, и поэтому $\bigcap\limits_{x \in M} A_x = \phi$, что противоречит построению A_x. Тем самым, $\bigcap\limits_{x \in C} A_x \neq \phi$. Следовательно, существует вектор $a \in \bigcap\limits_{x \in C} A_x$, для которого $a'x \geqslant 0 \ \forall x \in C$. Тогда в силу непрерывности скалярного произведения и $a'x \geqslant 0 \ \forall x \in \bar{C}$.

2. Применяя соответствующее линейное преобразование, мы можем добиться того, чтобы имело место соотношение

$$x \in L(C \cup \{0\}) \Leftrightarrow x = \begin{pmatrix} \hat{x} \\ 0 \end{pmatrix},$$

где $\hat{x} \in \mathbf{R}^r \subseteq \mathbf{R}^n$, $r = \dim L(C \cup \{0\})$.

Пусть $\hat{C} = \{\hat{x} \in \mathbf{R}^r : \begin{pmatrix} \hat{x} \\ 0 \end{pmatrix} \in C \subseteq \mathbf{R}^n\}$. Тогда, согласно доказанному выше в $\mathbf{R}^r$ существует гиперплоскость $\hat{E}(\hat{a}, 0)$ такая, что $\hat{a}'\hat{x} \geqslant 0 \ \forall \hat{x} \in \hat{C}$. Полагая $a = \begin{pmatrix} \hat{a} \\ 0 \end{pmatrix} \in \mathbf{R}^n$, получим $a'x \geqslant 0 \ \forall x \in C$

Очевидно, что хотя бы для одного $\hat{x} \in \hat{C}$ должно выполняться $\hat{a}'\hat{x} = a'x > 0$, поскольку иначе следовало бы $\mathbf{R}^r \supset \hat{E}(\hat{a}, 0) \supseteq \hat{C} \cup \{0\}$, что противоречит равенству $\dim L(C \cup \{0\}) = r$.

Теорема 2.34. (*первая теорема отделимости*). *Пусть* $C_1, C_2 \subseteq \mathbf{R}^n$ *– непустые выпуклые множества и* $\operatorname{ri} C_1 \cap \operatorname{ri} C_2 = \phi$. *Тогда существует гиперплоскость* $E(a, \alpha)$, *собственно разделяющая* $\bar{C}_1$ *и* $\bar{C}_2$.

Доказательство. Положим в теореме 2.33 $C = C_1 - C_2$. Согласно теореме 2.31 $\operatorname{ri} C = \operatorname{ri} C_1 - \operatorname{ri} C_2$, поэтому из теоремы 2.33 следует существование вектора $a \neq 0$ такого, что $a'z \geqslant 0 \ \forall z \in \overline{C_1 - C_2}$ и $a'\bar{z} > 0$ по крайней мере для одного $\bar{z} \in C_1 - C_2$. Полагая $z = x - y, x \in C_1, y \in C_2$, и вспоминая, что $\overline{C_1 - C_2} \supseteq \bar{C}_1 - \bar{C}_2$ (теорема 2.31), получим

$$a'x \geqslant a'y \quad \forall x \in \bar{C}_1 \quad \forall y \in \bar{C}_2, \tag{2.52}$$

$$a'\bar{x} > a'\bar{y} \quad \forall \bar{x} \in C_1 \quad \forall \bar{y} \in C_2. \tag{2.53}$$

При этом гиперплоскость $E(a, \alpha)$, где $\alpha = \inf\limits_{x \in C_1} a'x$, разделяет множества $\bar{C}_1$ и $\bar{C}_2$ собственно, потому что в силу (2.53) $E(a, \alpha) \not\supseteq C_1 \cup C_2$.

Замечание. Если в теореме 2.34 множество C_2 открыто, то гиперплоскость $E(a, \alpha)$ разделяет C_1 и C_2 строго. В самом деле, в этом слу-

чае с необходимостью выполняется неравенство $a'x^0 < \alpha \ \forall x^0 \in C_2$, потому что, с одной стороны, для достаточно малого $\epsilon > 0$ имеем $x^0 + \epsilon \frac{a}{\|a\|} \in C_2$, а с другой, из предположения $a'x^0 = \alpha$ следовало бы неравенство

$$a'\left(x^0 + \epsilon \frac{a}{\|a\|}\right) = a'x^0 + \epsilon\|a\| = \alpha + \epsilon\|a\| > \alpha,$$

что противоречит (2.52).

Следующее утверждение является следствием теоремы 2.34.

Теорема 2.35. *Пусть $V \subseteq \mathbf{R}^n$ – линейное многообразие, $C \subseteq \mathbf{R}^n$ – выпуклое множество и $V \cap \mathrm{ri}\, C = \phi$. Тогда существует гиперплоскость $E(a, \alpha)$, собственно разделяющая V и C и содержащая V.*

Доказательство. По определению $V = \mathrm{ri}\, V$, так что предположения теоремы 2.34 выполнены. Поэтому существует вектор $a \neq 0$ такой, что $a'x \geqslant a'y \ \forall x \in V, \ \forall y \in \bar{C}$, и $a'x > a'y$ хотя бы для одного $x \in V$ и одного $y \in \bar{C}$.

Без ограничения общности мы можем представить V в виде $V = \{x: x = b + Ay, \ y \in \mathbf{R}^m\}$, где A – матрица порядка $m \times n$, $b \in \mathbf{R}^n$. Полагая $\alpha = \inf\limits_{x \in V} a'x$, имеем $\alpha = \inf\limits_{y \in \mathbf{R}^m} (a'b + a'Ay)$. Для конечности α необходимо выполнение равенства $a'A = 0$. Отсюда следует $\alpha = a'b$ и окончательно

$$x \in V \Rightarrow a'x = a'(b + Ay) = \alpha \quad \forall y \in \mathbf{R}^m \Rightarrow x \in E(a, \alpha).$$

Теорема 2.36 (*вторая теорема отделимости*). *Пусть $C_1 \subseteq \mathbf{R}^n$ – непустое замкнутое выпуклое множество, $C_2 \subseteq \mathbf{R}^n$ – непустое компактное выпуклое множество и $C_1 \cap C_2 = \phi$. Тогда существует гиперплоскость $E(a, \alpha)$, строго разделяющая C_1 и C_2.*

Доказательство. Множество $-C_1$ выпукло и замкнуто; множество C_2 выпукло и компактно. Поэтому $C_2 - C_1$ выпукло и замкнуто. Очевидно, что $0 \notin C_2 - C_1$.

По теореме 2.32 существует гиперплоскость $E(a, \beta)$, $\beta > 0$, такая, что

$$a'(x - y) = a'x - a'y > \beta > 0 \quad \forall x \in C_2 \quad \forall y \in C_1.$$

Отсюда

$$\inf_{x \in C_2} a'x \geqslant \sup_{y \in C_1} a'y + \beta > \sup_{y \in C_1} a'y.$$

Если выбрать α из условия

$$\inf_{x \in C_2} a'x > \alpha > \sup_{y \in C_1} a'y,$$

то $E(a, \alpha)$ и будет искомой гиперплоскостью.

Замечание. Как показывает следующий пример, в теореме 2.36 требование компактности хотя бы одного из множеств C_1, C_2 является существенным.

Множества

$$C_1 = \{(x_1, x_2)' : x_2 = 0\},$$

$$C_2 = \{(x_1, x_2)' : x_1, x_2 \in \operatorname{int} \mathbf{R}_+,\ x_2 \geqslant 1/x_1\}$$

замкнуты и выпуклы, но не компактны. Очевидно, что не существует гиперплоскости (прямой), которая бы строго разделяла C_1 и C_2. Поскольку оба множества удовлетворяют условиям первой теоремы отделимости, найдется гиперплоскость (прямая $E(a, 0) = C_1$), разделяющая C_1 и C_2.

Как следствие из второй теоремы отделимости получается следующее утверждение.

Теорема 2.37. *Любое замкнутое выпуклое множество $C \subsetneq \mathbf{R}^n$ есть пересечение всех замкнутых полупространств F, содержащих C.*

Доказательство. Пусть D — пересечение всех замкнутых полупространств F таких, что $F \supseteq C$. Так как, очевидно, $C \subseteq D$, то остается лишь показать, что $C \supseteq D$. Пусть $x^0 \in D \backslash C$. Положим $C_1 = C$, $C_2 = \{x^0\}$. С помощью простых рассуждений из теоремы 2.35 заключаем, что существует гиперплоскость $E(a, \alpha)$, для которой

$$a'x^0 < \alpha < a'x \quad \forall x \in C_1.$$

Отсюда имеем

$$C \subseteq E_+(a, \alpha) = \{x \in \mathbf{R}^n : a'x \geqslant \alpha\}, \quad x^0 \notin E_+(a, \alpha)$$

и, следовательно, $x^0 \notin D$, $D \backslash C = \phi$.

Из теоремы 2.36 вытекает также следующий результат.

Теорема 2.38. *Любое замкнутое выпуклое множество $C \subsetneq \mathbf{R}^n$ и любая точка $x \notin C$ могут быть строго разделены некоторой гиперплоскостью $E(a, \alpha)$.*

Для произвольных множеств имеет место следующее утверждение, соответствующее теореме 2.37.

Теорема 2.39. *Пусть $M \subsetneq \mathbf{R}^n$ — произвольное множество. Тогда $\overline{C(M)}$ есть пересечение всех замкнутых полупространств F, содержащих M.*

Доказательство. Для замкнутого полупространства F, очевидно, выполняется условие

$$F \supseteq \overline{C(M)} \Leftrightarrow F \supseteq M.$$

В качестве следствия мы докажем одно утверждение о выпуклых замкнутых конусах, из которого вытекает известная теорема Фаркаша (теорема 2.43).

Теорема 2.40. *Пусть $M \subsetneq \mathbf{R}^n$ — непустой замкнутый выпуклый конус (т.е. $M = \overline{K^c(M)} \neq \phi$). Тогда M есть пересечение всех однородных замкнутых полупространств $E_+(a, 0)$, содержащих M.*

Доказательство. Требуется показать, что

$$a'x \geqslant \alpha \quad \forall x \in M \Rightarrow a'x \geqslant 0 \quad \forall x \in M.$$

Очевидно, в силу $0 \in M$ должно выполняться $\alpha \leqslant 0$. Допустим, существует вектор $x^0 \in M$, для которого $a'x^0 < 0$. Тогда для $\lambda x^0 \in M$ и $\lambda > 0$ имеем $a'x^0\lambda < 0$, так что скалярное произведение $a'x$ оказывается неограниченным снизу на M. Поскольку этого быть не может, то получается $a'x \geqslant 0$ $\forall x \in M$.

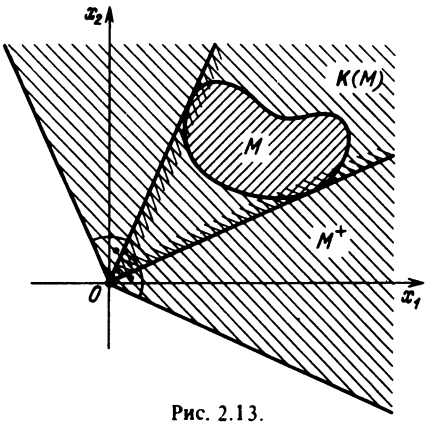


Рис. 2.13.

Обобщая ситуацию, описанную в теореме 2.40, рассмотрим совокупность всех однородных замкнутых полупространств $E_+(a, 0)$, содержащих произвольное множество M. Эту совокупность мы охарактеризуем с помощью множества M^+ всех принадлежащих $E_+(a, 0)$ векторов-нормалей a:

$$M^+ = \{a: a'x \geqslant 0 \ \ \forall x \in M\}.$$

Очевидно, что M^+ – выпуклый замкнутый конус. Назовем M^+ *полярным конусом* к множеству M (рис. 2.13).

Пример 2.11. Для выпуклого конуса M, задаваемого условием

$M = \{x: x = Ay, y \in \mathbf{R}^m_+; A$ – матрица порядка $n \times m\}$,

имеем

$$M^+ = \{a: a'x \geqslant 0 \ \ \forall x \in M\} = \{a: a'Ay \geqslant 0 \ \ \forall y \in \mathbf{R}^m_+\} = \{a: a'A \geqslant 0\}.$$

Таким образом, M^+ – многогранный конус.

В следующей теореме содержатся некоторые легко проверяемые элементарные свойства полярных конусов.

Теорема 2.41. *Пусть M_1 и M_2 – множества из $\mathbf{R}^n$. Справедливы соотношения*:

$$1)\ M_1 \supseteq M_2 \Rightarrow M_1^+ \subseteq M_2^+, \tag{2.54}$$

$$2)\ (M_1 \cup M_2)^+ = M_1^+ \cap M_2^+. \tag{2.55}$$

Если к M^+ применить еще раз операцию образования полярного конуса, то получится множество $M^{++} = (M^+)^+$. Рассмотрим, каким образом M^{++} связано с M. Прежде всего, по определению M^+, имеем

$$a \in M^+ \Rightarrow \{x: a'x \geqslant 0\} \supseteq M.$$

Далее,

$$M^{++} = \{x: x'a \geqslant 0 \ \ \forall a \in M^+\} = \bigcap_{a \in M^+} \{x: a'x \geqslant 0\},$$

т.е. M^{++} – пересечение всех полупространств $E_+(a, 0)$, содержащих M. Следовательно, для $M \subseteq \mathbf{R}^n$ выполняется

$$M^{++} \supseteq M. \tag{2.56}$$

С учетом первого утверждения теоремы 2.41 далее получаем

$$M^+ = M^{+++}, \ M^{++} = M^{++++}, \ldots, \tag{2.57}$$

поскольку, с одной стороны, имеем

$$M \subseteq M^{++} \Rightarrow M^+ \supseteq (M^{++})^+ = M^{+++},$$

а с другой, из (2.56) при подстановке M^+ вместо M вытекает

$$M^+ \subseteq (M^+)^{++} = M^{+++}.$$

Из соотношений (2.54), (2.56) и (2.57) получаем, что соответствие между M^{++} и M является операцией взятия оболочки.

Следующая теорема описывает множества, которые характеризуются этой операцией.

Теорема 2.42. *Непустое множество $M \subseteq \mathbf{R}^n$ тогда и только тогда является замкнутым выпуклым конусом, когда выполняется равенство $M^{++} = M$.*

Доказательство. 1. $M^{++} = M$ означает, что M есть пересечение всех однородных замкнутых полупространств, содержащих M. Тогда очевидно, что M – замкнутый выпуклый конус.

2. Пусть M – замкнутый выпуклый конус, $M \neq \phi$. Тогда по теореме 2.40 M есть пересечение всех однородных замкнутых полупространств, содержащих M, и, следовательно, $M = M^{++}$.

Так как положительная оболочка конечного числа векторов является замкнутым выпуклым конусом, то, как следствие предыдущего результата, получаем следующую теорему.

Теорема 2.43 (*теорема Фаркаша*). *Пусть B – матрица порядка $m \times n$ и $b \in \mathbf{R}^n$. Тогда из двух систем*:

$$S_1\colon\ Bx \leqslant 0,\ b'x > 0,\ x \in \mathbf{R}^n,$$

$$S_2\colon\ v'B = b',\ v \geqslant 0,\ v \in \mathbf{R}^m,$$

лишь одна имеет решение.

Доказательство. По теореме 2.17 множество $M = \{u \in \mathbf{R}^n\colon u = B'v,\ v \geqslant 0\}$ – замкнутый выпуклый конус. Согласно теореме 2.42 имеем

$$b \in M \Leftrightarrow b'x \geqslant 0 \quad \forall x \in M^+ = \{x\colon Bx \geqslant 0\},$$

следовательно,

$$b \in M \Leftrightarrow b'x \leqslant 0 \quad \forall x \in -M^+ = \{x\colon Bx \leqslant 0\}.$$

Утверждение теоремы 2.43 может быть проинтерпретировано и геометрически. Пусть S_1 и S_2 представлены в виде

$$S_1\colon\quad {}^{j}bx \leqslant 0,\quad j = 1, \ldots, m,\quad b'x > 0,$$

$$S_2\colon\quad \sum_{j=1}^{m} v_j ({}^{j}b)' = b,\qquad v_j \geqslant 0,\qquad j = 1, \ldots, m.$$

Система S_1 требует существования такого вектора $x \in \mathbf{R}^n$, который образует со всеми векторами $({}^1b)', \ldots, ({}^m b)'$ тупой угол ($\geqslant \pi/2$), а с b – острый угол ($< \pi/2$). Напротив, S_2 требует, чтобы вектор b был представим в виде неотрицательной линейной комбинации векторов $({}^j b)'$. Для случая $n = 2, m = 3$ система S_1 на рис. 2.14,*а* и система S_2 на рис. 2.14,*б* не имеют решения.

Из теорем 2.43 и 2.11 вытекает следующее утверждение.

Теорема 2.44 (*теорема Вейля*). *Положительная оболочка конечного числа векторов есть многогранный конус*

Доказательство. Положительная оболочка векторов $a^1, a^2, \ldots, a^n$ – это множество $K = \{y\colon y = Ax,\ x \geqslant 0\}$, где $A = (a^1, \ldots, a^n)$. Полярный конус $K^+ = \{u\colon u'A \geqslant 0\}$ является многогранным конусом (см.

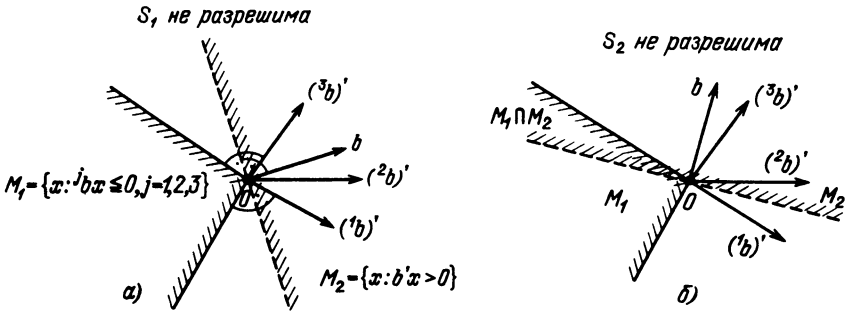


Рис. 2.14.

пример 2.11). Поэтому, согласно теореме 2.11, существует матрица B такая, что

$$K^+ = \{u:\ u = Bv,\ v \geqslant 0\}.$$

Далее, $K^{++} = \{y:\ y'B \geqslant 0\}$, и в силу теоремы Фаркаша справедливо равенство

$$K = K^{++} = \{y:\ y'B \geqslant 0\}.$$

2.1.5. Экстремальные точки и опорные гиперплоскости. Введем понятие экстремальной точки выпуклого множества. Точка x выпуклого множества C называется *экстремальной* точкой C, если не существует двух точек $x^1,\ x^2 \in C$ таких, что

$$x = \lambda_1 x^1 + \lambda_2 x^2;\quad x^1 \neq x^2;\quad (\lambda_1, \lambda_2)' \in \mathrm{ri}\, P_+^2.$$

Экстремальные точки характеризует следующая теорема.

Т е о р е м а 2.45. *Точка x тогда и только тогда является экстремальной точкой выпуклого множества C, когда не существует двух точек $x^1,\ x^2 \in C$, удовлетворяющих соотношению*

$$x = (x^1 + x^2)/2,\quad x^1 \neq x^2. \tag{2.58}$$

Д о к а з а т е л ь с т в о. Так как необходимость условия (2.58) видна непосредственно, мы ограничимся доказательством достаточности, которое проведем от противного.

Допустим, что для некоторого вектора $x \in C$ не существует представления вида (2.58), но он может быть выражен в виде $x = \lambda_1 x^1 + \lambda_2 x^2$, $(\lambda_1, \lambda_2)' \in \mathrm{ri}\, P_+^2$. Выберем теперь $\epsilon > 0$ настолько малым, чтобы числа $\alpha = \lambda_1 + \epsilon$ и $\beta = \lambda_1 - \epsilon$ удовлетворяли условию $\alpha, \beta \in [0, 1]$. Поскольку C выпукло, то

$$x^1, x^2 \in C \Rightarrow \begin{cases} a = \alpha x^1 + (1-\alpha)x^2 \in C, \\ b = \beta x^1 + (1-\beta)x^2 \in C \end{cases}$$

и, далее,

$$x = \lambda_1 x^1 + \lambda_2 x^2 = \lambda_1 x^1 + (1 - \lambda_1)x^2 = (a + b)/2,$$

что противоречит исходному предположению. Отсюда x – экстремальная точка C.

Пример 2.12. Экстремальные точки сегмента $[x^1, x^2]$ – его концевые точки x^1, x^2. Экстремальные точки треугольника – три точки, на которые он натянут. Вообще, экстремальные точки симплекса совпадают с его вершинами.

Пример 2.13. Пусть A – матрица порядка $m \times n$ и $b \in \mathbf{R}^m$. Как известно из теории линейного программирования (см. [84]), вектор $x^0 \in G = \{x \in \mathbf{R}^n: Ax = b\}$ тогда и только тогда является экстремальной точкой выпуклого множества G, когда соответствующие положительным координатам x^0 столбцы матрицы A линейно независимы.

Пример 2.14. Для замкнутого шара $\bar{S}\left(\begin{pmatrix}0\\0\end{pmatrix}, 1\right) \subset \mathbf{R}^2$ все точки границы $\bar{S}\left(\begin{pmatrix}0\\0\end{pmatrix}, 1\right) \setminus S\left(\begin{pmatrix}0\\0\end{pmatrix}, 1\right)$ являются экстремальными точками. Открытое же множество $S\left(\begin{pmatrix}0\\0\end{pmatrix}, 1\right)$ не имеет экстремальных точек.

Теорема 2.46. *Пусть C – выпуклое множество и $x \in C$. Тогда справедливы соотношения*:

1) *x – экстремальная точка $C \Rightarrow x$ – относительная граничная точка C*;

2) *x – экстремальная точка $C \Leftrightarrow C \setminus \{x\}$ выпукло.*

Доказательство. 1. Без ограничения общности можно считать, что $\dim C = n$. Допустим, что экстремальная точка $x \in C$ одновременно является внутренней точкой C. Тогда существует окрестность $U(x) = \{y: \|y - x\| < \epsilon\} \subseteq C$. Если $x^1 \in U(x)$ и $x^2 = -x^1 + 2x$, то $\|x^2 - x\| = \|x^1 - x\| < \epsilon$, т.е. $x^2 \in U(x)$, и в то же время $x = (x^1 + x^2)/2$, что противоречит выбору x. Следовательно, x – граничная точка C.

2. а) Пусть множество $C \setminus \{x\}$ не выпукло. Тогда найдутся две точки $x^1, x^2 \in C \setminus \{x\}$ такие, что отрезок $[x^1, x^2]$ не лежит полностью в $C \setminus \{x\}$. Но поскольку $[x^1, x^2] \in C$, а C отличается от $C \setminus \{x\}$ лишь добавлением точки x, должно выполняться $x \in [x^1, x^2]$. Очевидно, $x^1 \neq x$, $x^2 \neq x$, так что x не может быть экстремальной точкой C.

б) Пусть x не является экстремальной точкой C. Тогда найдутся точки $x^1, x^2 \in C$, $x^1 \neq x$, $x^2 \neq x$, такие, что $x \in [x^1, x^2]$. Очевидно, $x^1, x^2 \in C \setminus \{x\}$, в то время как $[x^1, x^2]$ не лежит полностью в $C \setminus \{x\}$. Поэтому множество $C \setminus \{x\}$ не может быть выпукло.

Нетрудно понять, что первое утверждение теоремы 2.46 не обратимо.

Как мы уже установили, открытое выпуклое множество не содержит экстремальных точек (см. пример 2.14). Более того, имеются также замкнутые выпуклые множества, например, замкнутое полупространство, которые не обладают экстремальными точками. Следующая теорема дает достаточное условие для существования экстремальной точки.

Теорема 2.47. *Любое непустое компактное выпуклое множество $C \subseteq \mathbf{R}^n$ имеет хотя бы одну экстремальную точку.*

Доказательство. В силу компактности C существует вектор $x^0 \in C$ такой, что

$$\|x^0\| \geqslant \|x\| \quad \forall x \in C. \tag{2.59}$$

Предположим, что существует представление

$$x^0 = (x^1 + x^2)/2, \quad x^1, x^2 \in C, \tag{2.60}$$

и покажем, что отсюда следует $x^1 = x^2$. Тогда по теореме 2.45 x^0 должен быть экстремальной точкой C.

Из (2.60) вытекает $x^0 = x^1 + (x^2 - x^1)/2$ и, в силу (2.59),

$$\| x^1 \|^2 \leqslant \| x^0 \|^2 = \| x^1 \|^2 + (x^2 - x^1)' x^1 + \| x^2 - x^1 \|^2/4,$$

или

$$\| x^2 - x^1 \|^2 \geqslant -4(x^2 - x^1)' x^1. \qquad (2.61)$$

Аналогично из (2.60) следует $x^0 = x^2 + (x^1 - x^2)/2$, и поэтому $\| x^1 - x^2 \|^2 \geqslant 4(x^2 - x^1)' x^2$. Сложение последнего неравенства с (2.61) дает $0 \geqslant 2\| x^2 - x^1 \|^2$ и, тем самым, $x^1 = x^2$. Следовательно, x^0 – экстремальная точка C.

Замечание. Как показывают простые примеры, ни замкнутость, ни ограниченность множества C не могут быть опущены из условий теоремы 2.47.

Теорема 2.48 (*теорема Крейна – Мильмана*). *Любое непустое компактное выпуклое множество C совпадает с замыканием выпуклой оболочки своих экстремальных точек.*

Доказательство. Обозначим через $\ddot{C}$ множество экстремальных точек C. Справедливы соотношения

C выпукло $\Rightarrow C(\ddot{C}) \subseteq C$,

C замкнуто $\Rightarrow \overline{C(\ddot{C})} \subseteq C$.

Таким образом, остается доказать $\overline{C(\ddot{C})} \supseteq C$, что мы и сделаем от противного. Допустим, найдется $x^0 \in C$, $x^0 \notin \overline{C(\ddot{C})}$. По теореме 2.36 существует гиперплоскость $E(a, \alpha)$ такая, что $a'x > a'x^0 \ \forall x \in \overline{C(\ddot{C})}$, и, тем самым, $a'x > a'x^0 \ \forall x \in \ddot{C}$. Последнее же неравенство противоречиво, так как линейная функция $a'x$ достигает своего минимума на C хотя бы в одной экстремальной точке C, т.е. в точке множества $\ddot{C}$. Следовательно, $C = \overline{C(\ddot{C})}$.

Замечание. Условия теоремы 2.48 существенны для полученного утверждения. Так, например, неограниченный многогранный выпуклый конус не совпадает с замыканием выпуклой оболочки своих экстремальных точек.

Ряд утверждений об экстремальных точках выпуклых множеств мы получим, определив понятие опорной гиперплоскости. *Гиперплоскость* $E(a, \alpha)$ называется *опорной* к непустому (не обязательно выпуклому) множеству M, если

$$\inf_{x \in M} a'x = \alpha.$$

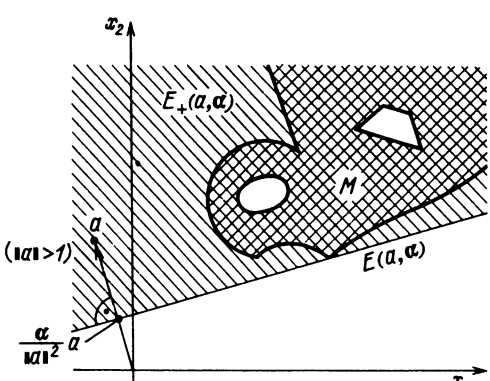


Рис. 2.15.

Точка $\hat{x} \in M \cap E(a, \alpha)$ называется *опорной точкой* опорной гиперплоскости $E(a, \alpha)$.

Имеет место неравенство $a'x \geqslant \alpha \ \forall x \in M$, т.е. все точки M лежат в полупространстве $E_+(a, \alpha) = \{x: a'x \geqslant \alpha\}$, которое называется *опорным полу-*

пространством множества M. На рис. 2.15 наряду с вектором нормали a для $E(a, \alpha)$ отмечена также ближайшая к началу координат точка $\frac{\alpha}{\|a\|^2} a$ этой гиперплоскости, что поясняет геометрический смысл числа α.

Если M – компактное множество, то для любого $a \in \mathbf{R}^n$, $a \neq 0$, существует опорная гиперплоскость к M. Для этого нужно лишь положить $\alpha = \min\limits_{x \in M} a'x$.

Если опорная гиперплоскость $E(a, \alpha)$ к множеству M содержит само M, то мы говорим о *собственной опорной гиперплоскости*, в противном случае – о *несобственной опорной гиперплоскости.*

Для выпуклых множеств имеет место следующее утверждение.

Теорема 2.49. *Пусть $C \subseteq \mathbf{R}^n$ – выпуклое множество. Тогда*:

1) *любая опорная гиперплоскость $E(a, \alpha)$ к множеству C, содержащая точку $x^0 \in \mathrm{ri}\, C$, является собственной*;

2) *через любую относительную граничную точку $x^0 \in C$, если только C не состоит из одной точки, проходит по меньшей мере одна несобственная опорная гиперплоскость к множеству C.*

Доказательство. 1. Для опорной гиперплоскости $E(a, \alpha)$ выполняется $a'x \geqslant \alpha \quad \forall x \in C$. Аффинную оболочку C можно представить в виде

$$V(C) = \{x:\ x = x^0 + Ay, \quad y \in \mathbf{R}^m\},$$

где $x^0 \in \mathrm{ri}\, C$. По предположению $a'x^0 = \alpha$, и в силу $x^0 \in \mathrm{ri}\, C$ для достаточно малого $\lambda > 0$ имеем

$$a'(x^0 + Ay) \geqslant \alpha \quad \forall y \in S(0, \lambda).$$

Отсюда следует $a'Ay \geqslant 0 \quad \forall y \in S(0, \lambda)$, и $a'A = 0$.

Так как $a'(x^0 + Ay) = a'x^0 = \alpha \quad \forall y \in \mathbf{R}^n$, то $E(a, \alpha) \supseteq V(C) \supseteq C$.

2. Полагая в теореме 2.34 $C_1 = C$ и $C_2 = \{x^0\}$, получим доказываемое утверждение.

Теорема 2.50. *Пусть C – непустое компактное выпуклое множество. Тогда любая опорная гиперплоскость к множеству C содержит хотя бы одну экстремальную точку C.*

Доказательство. Если $E(a, \alpha)$ – произвольная опорная гиперплоскость к множеству C, то $D = E(a, \alpha) \cap C$ – непустое компактное выпуклое множество, которое по теореме 2.47 имеет хотя бы одну экстремальную точку x^0.

Покажем теперь, что x^0 является также экстремальной точкой множества C. Для этого допустим, что существует представление $x^0 = (x^1 + x^2)/2$, $x^1, x^2 \in C$. По определению $E(a, \alpha)$ выполняется $a'x^1 \geqslant \alpha$, $a'x^2 \geqslant \alpha$. В силу $x^0 \in E(a, \alpha)$ имеем $a'x^0 = \alpha$. Поэтому из равенства

$$a'\left(\frac{1}{2}\left(x^1 + x^2\right)\right) = \frac{1}{2}a'x^1 + \frac{1}{2}a'x^2 = a'x^0 = \alpha$$

следует $a'x^1 = a'x^2 = \alpha$ и, тем самым, $x^1, x^2 \in D = E(a, \alpha) \cap C$.

Поскольку x^0 – экстремальная точка D, должно выполняться $x^1 = x^2$. Тогда по теореме 2.44 x^0 – экстремальная точка C.

2.2. Выпуклые функции

2.2.1. Определение и основные свойства выпуклых функций. Класс выпуклых (и вогнутых) функций играет в нелинейном программировании важную роль. Это объясняется среди прочего тем, что у выпуклой функции каждый локальный минимум одновременно является и глобальным.

Определенная на непустом выпуклом множестве $D(f) \subseteq \mathbf{R}^n$ функция f. $D(f) \to \mathbf{R}$ называется *выпуклой* на $D(f)$, если выполняется соотношение

$$\left.\begin{array}{l} x^1, x^2 \in D(f) \\ (\lambda_1, \lambda_2)' \in P_+^2 \end{array}\right\} \Rightarrow f(\lambda_1 x^1 + \lambda_2 x^2) \leqslant \lambda_1 f(x^1) + \lambda_2 f(x^2).$$

Функция f называется *вогнутой* на $D(f)$, если $-f$ выпукла на $D(f)$. Функция f называется *строго выпуклой* на $D(f)$, если

$$\left.\begin{array}{l} x^1, x^2 \in D(f),\ x^1 \neq x^2 \\ (\lambda_1, \lambda_2)' \in \operatorname{ri} P_+^2 \end{array}\right\} \Rightarrow f(\lambda_1 x^1 + \lambda_2 x^2) < \lambda_1 f(x^1) + \lambda_2 f(x^2).$$

Функция f называется *строго вогнутой* на $D(f)$, если $-f$ строго выпукла на $D(f)$.

Вектор-функция f: $D(f) \to \mathbf{R}^m$, $D(f) \subseteq \mathbf{R}^n$, где $f = (f_1, ..., f_m)'$, называется выпуклой, вогнутой, строго выпуклой или строго вогнутой, если каждая из функций-координат $f_1, ..., f_m$ обладает соответствующими свойствами.

Следующие ниже утверждения содержат простые примеры выпуклых и строго выпуклых функций. При этом в силу тесной связи между выпуклыми и вогнутыми функциями мы ограничимся в большинстве случаев выпуклыми функциями.

Т е о р е м а 2.51. *Функция f: $\mathbf{R}^n \to \mathbf{R}$ тогда и только тогда является аффинной, когда она одновременно выпукла и вогнута.*

Д о к а з а т е л ь с т в о. 1. То, что каждая аффинная функция f: $\mathbf{R}^n \to \mathbf{R}$ вида $f(x) = a'x + b (a \in \mathbf{R}^n,\ b \in \mathbf{R})$ является как выпуклой, так и вогнутой, ясно непосредственно.

2. Пусть теперь f и выпукла, и вогнута. Тогда для произвольного фиксированного вектора $(\lambda_1, \lambda_2)' \in P_+^2$ имеем

$$f(\lambda_1 x^1 + \lambda_2 x^2) = \lambda_1 f(x^1) + \lambda_2 f(x^2) \quad \forall x^1, x^2 \in \mathbf{R}^n.$$

Следовательно, для функции $g(x) = f(x) - f(0)$ и произвольного $(\lambda_1, \lambda_2)' \in P_+^2$ выполняется равенство

$$g(\lambda_1 x^1 + \lambda_2 x^2) = \lambda_1 g(x^1) + \lambda_2 g(x^2) \quad \forall x^1, x^2 \in \mathbf{R}^n.$$

Кроме того, очевидно, $g(0) = 0$.

Пусть $x \in \mathbf{R}^n$ – произвольный вектор. Тогда для $\mu \in [0, 1]$ получим

$$g(\mu x) = g((1 - \mu)0 + \mu x) = (1 - \mu) g(0) + \mu g(x) = \mu g(x).$$

Для $\mu > 1$ будем иметь

$$g(x) = g\left(\left(1 - \frac{1}{\mu}\right)0 + \frac{1}{\mu}(\mu x)\right) = \frac{1}{\mu} g(\mu x).$$

Для $\mu < 0$ с учетом равенств

$$0 = g(0) = g\left(\frac{1}{2}(-x) + \frac{1}{2}x\right) = \frac{1}{2}g(-x) + \frac{1}{2}g(x)$$

получим соотношение

$$g(\mu x) = g((-\mu)(-x)) = -\mu g(-x) = \mu g(x).$$

Таким образом, для произвольного $x \in \mathbf{R}^n$ и всех $\mu \in \mathbf{R}$ выполняется равенство $g(\mu x) = \mu g(x)$. Поэтому

$$g(x+y) = g\left(\frac{1}{2}2x + \frac{1}{2}2y\right) = \frac{1}{2}g(2x) + \frac{1}{2}g(2y) = g(x) + g(y),$$

т.е. g: $\mathbf{R}^n \to \mathbf{R}$ – линейная функция и представима в виде $g(x) = a'x$ для некоторого $a \in \mathbf{R}^n$. Отсюда $f(x) = a'x + b$, где $b := f(0)$.

Пример 2.15. Функция $f(\mu) = \mu^2$, $\mu \in \mathbf{R}$, строго выпукла, потому что для $\mu_1, \mu_2 \in \mathbf{R}$, $\mu_1 \neq \mu_2$, и $\lambda \in (0, 1)$ выполняется

$$f((1-\lambda)\mu_1 + \lambda\mu_2) = f(\mu_1 + \lambda(\mu_2 - \mu_1)) = \mu_1^2 + 2\lambda\mu_1(\mu_2 - \mu_1) +$$
$$+ \lambda^2(\mu_2 - \mu_1)^2 < \mu_1^2 + 2\lambda\mu_1(\mu_2 - \mu_1) + \lambda(\mu_2 - \mu_1)^2 = (1-\lambda)f(\mu_1) + \lambda f(\mu_2).$$

Теорема 2.52. *Пусть $D(f) \subseteq \mathbf{R}^n$ – выпуклое множество. Если функции $f_1, \ldots, f_m$ выпуклы на $D(f)$ и $\alpha_1, \ldots, \alpha_m \in \mathbf{R}_+$, то функция $f = \sum_{i=1}^{m} \alpha_i f_i$ также выпукла на $D(f)$.*

Доказательство. Пусть x^1, $x^2 \in D(f)$ и $(\lambda_1, \lambda_2)' \in P_+^2$. Тогда

$$\sum_{i=1}^{m} \alpha_i f_i(\lambda_1 x^2 + \lambda_2 x^2) \leqslant \lambda_1 \sum_{i=1}^{m} \alpha_i f_i(x^1) + \lambda_2 \sum_{i=1}^{m} \alpha_i f_i(x^2),$$

т.е.

$$f(\lambda_1 x^1 + \lambda_2 x^2) \leqslant \lambda_1 f(x^1) + \lambda_2 f(x^2).$$

Сумма строго выпуклой и выпуклой функций является строго выпуклой функцией. Например, $f(\mu) = \alpha\mu^2 + \beta\mu + \gamma$ при $\alpha > 0$ строго выпукла, а при $\alpha = 0$ выпукла.

Теорема 2.53. *Пусть $D(f) \subseteq \mathbf{R}^n$ – выпуклое множество. Функция f тогда и только тогда выпукла на $D(f)$, когда для произвольных фиксированных $x, y \in D(f)$ функция $\varphi(\lambda) = f(\lambda x + (1-\lambda)y)$ выпукла на сегменте $[0, 1]$.*

Доказательство. 1. Пусть $\varphi(\lambda)$ выпукла на $[0, 1]$. Для произвольных фиксированных $x, y \in D(f)$ и любых $(\lambda_1, \lambda_2)' \in P_+^2$ имеем

$$f(\lambda_1 x + \lambda_2 y) = \varphi(\lambda_1 \cdot 1 + \lambda_2 \cdot 0) \leqslant \lambda_1\varphi(1) + \lambda_2\varphi(0) = \lambda_1 f(x) + \lambda_2 f(y).$$

2. Обратно, пусть f выпукла на $D(f)$. Для произвольных фиксированных $x, y \in D(f)$ и $\lambda_1, \lambda_2 \in [0, 1]$, а также для любых $(\mu_1, \mu_2)' \in P_+^2$

выполняется

$$\varphi(\mu_1\lambda_1 + \mu_2\lambda_2) = f((\mu_1\lambda_1 + \mu_2\lambda_2)x + (1 - \mu_1\lambda_1 - \mu_2\lambda_2)y) =$$

$$= f(\mu_1(\lambda_1 x + (1 - \lambda_1)y) + \mu_2(\lambda_2 x + (1 - \lambda_2)y)) \leqslant$$

$$\leqslant \mu_1 f(\lambda_1 x + (1 - \lambda_1)y) + \mu_2 f(\lambda_2 x + (1 - \lambda_2)y) = \mu_1\varphi(\lambda_1) + \mu_2\varphi(\lambda_2).$$

В следующем утверждении речь идет о свойствах выпуклости квадратичной формы.

Т е о р е м а 2.54. *Пусть A – симметрическая матрица порядка $n \times n$ и Q – квадратичная форма $Q(x) = x'Ax$, $x \in \mathbf{R}^n$. Тогда*:

1) *из положительной полуопределенности $Q(x)$ следует выпуклость $Q(x)$*;

2) *из положительной определенности $Q(x)$ следует строгая выпуклость $Q(x)$*.

Д о к а з а т е л ь с т в о. 1. Пусть $Q(x)$ положительно полуопределена. Тогда для произвольных $x, y \in \mathbf{R}^n$ и $\lambda \in [0, 1]$ имеем

$$Q(\lambda x + (1 - \lambda)y) =$$

$$= x'Ay + 2(1 - \lambda)(y - x)'Ax + (1 - \lambda)^2(y - x)'A(y - x).$$

В силу неравенства $(y - x)'A(y - x) \geqslant 0 \quad \forall x, y \in \mathbf{R}^n$ выполняется соотношение

$$Q(\lambda x + (1 - \lambda)y) \leqslant x'Ax + 2(1 - \lambda)(y - x)'Ax + (1 - \lambda)(y - x)'A(y - x) =$$

$$= x'Ax + (1 - \lambda)(y - x)'Ax + (1 - \lambda)(y - x)'Ay = \lambda Q(x) + (1 - \lambda)Q(y).$$

2. Второе утверждение доказывается аналогично первому.

З а м е ч а н и е. Соответствующие утверждения могут быть сформулированы для отрицательно полуопределенных и отрицательно определенных квадратичных форм.

Т е о р е м а 2.55. *Пусть $D(f) \subseteq \mathbf{R}^n$ – выпуклое множество и $f\colon D(f) \to \mathbf{R}^m$ – выпуклая вектор-функция. Тогда для любого $a \in \mathbf{R}^m$ множества уровня $N_a^0 = \{x \in D(f)\colon f(x) < a\} \subseteq D(f)$ и $N_a = \{x \in D(f)\colon f(x) \leqslant a\} \subseteq D(f)$ выпуклы.*

Д о к а з а т е л ь с т в о. Ограничимся случаем строгого неравенства. Пусть задан произвольный фиксированный вектор $a \in \mathbf{R}^m$. Для произвольных точек $x^1, x^2 \in N_a^0$ и произвольного вектора $(\lambda_1, \lambda_2)' \in P_+^2$ имеем

$$f(\lambda_1 x^1 + \lambda_2 x^2) \leqslant \lambda_1 f(x^1) + \lambda_2 f(x^2) < a;$$

таким образом, множество N_a^0 выпукло.

З а м е ч а н и е. Утверждение, обратное к теореме 2.55, не имеет места. Это показывает следующий пример:

$$f(x) = \operatorname{tg} x; \quad D(f) = (-\pi/2, \pi/2) \subseteq \mathbf{R}$$

(сравните со свойствами непрерывных квазивыпуклых функций из теоремы 2.93).

Для того чтобы при рассмотрении выпуклых функций можно было использовать свойства выпуклых множеств, мы введем для произвольной функции $f\colon D(f) \to \mathbf{R}$ следующие множества:

надграфик (или *эпиграф*) f:

$$\operatorname{epi} f = \left\{ \begin{pmatrix} x \\ \xi \end{pmatrix} \in \mathbf{R}^{n+1} : \; f(x) \leqslant \xi, \; x \in D(f) \right\},$$

подграфик (или *гипограф*) f [*)]:

$$\operatorname{hypo} f = \left\{ \begin{pmatrix} x \\ \xi \end{pmatrix} \in \mathbf{R}^{n+1} : \; f(x) \geqslant \xi, \; x \in D(f) \right\}.$$

Соответствующие множества можно определить и для векторных функций $f: D(f) \to \mathbf{R}^m$.

На рис. 2.16,*а* изображен надграфик выпуклой функции $f_1: D(f_1) \to \mathbf{R}$, на рис. 2.16,*б* – подграфик вогнутой функции $f_2: D(f_2) \to \mathbf{R}$. Видно, что $\operatorname{epi} f_1$ и $\operatorname{hypo} f_2$ являются выпуклыми множествами в $\mathbf{R}^2$. Справедливо и обратное утверждение.

Т е о р е м а 2.56. *Функция* $f: D(f) \to \mathbf{R}$ *тогда и только тогда выпуклая, когда* $\operatorname{epi} f$ *– выпуклое множество.*

Д о к а з а т е л ь с т в о. 1. Пусть f выпукла. Для произвольных $(x^1, \xi^1)'$, $(x^2, \xi^2)' \in \operatorname{epi} f$ и $(\lambda_1, \lambda_2)' \in P_+^2$ имеем

$$f(\lambda_1 x^1 + \lambda_2 x^2) \leqslant \lambda_1 f(x^1) + \lambda_2 f(x^2) \leqslant \lambda_1 \xi^1 + \lambda_2 \xi^2.$$

Отсюда следует, что $\begin{pmatrix} \lambda_1 x^1 + \lambda_2 x^2 \\ \lambda_1 \xi^1 + \lambda_2 \xi^2 \end{pmatrix} \in \operatorname{epi} f$, т.е. $\operatorname{epi} f$ – выпуклое множество.

2. Пусть $\operatorname{epi} f$ – выпуклое множество. Допустим, что f не выпукла. Тогда существуют точки $x^1, x^2 \in D(f)$, $x^1 \neq x^2$, и $(\lambda_1^*, \lambda_2^*)' \in P_+^2$ такие, что

$$f(\lambda_1^* x^1 + \lambda_2^* x^2) > \lambda_1^* f(x^1) + \lambda_2^* f(x^2). \tag{2.62}$$

[*)] Другие названия: *верхний*, соответственно *нижний полуцилиндры* функции f.

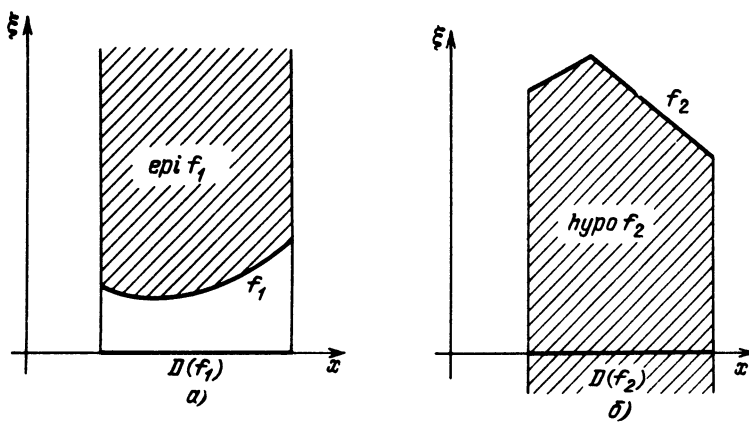


Рис. 2.16.

С одной стороны, всегда $\begin{pmatrix} x^1 \\ f(x^1) \end{pmatrix}, \begin{pmatrix} x^2 \\ f(x^2) \end{pmatrix} \in \operatorname{epi} f$, а с другой, в силу (2.62), имеем

$$\begin{pmatrix} \lambda_1^* x^1 + \lambda_2^* x^2 \\ \lambda_1^* f(x^1) + \lambda_2^* f(x^2) \end{pmatrix} \notin \operatorname{epi} f,$$

т.е. множество $\operatorname{epi} f$ не выпукло, что противоречит исходному предположению.

Некоторые топологические свойства надграфика функции мы рассмотрим в 2.2.6.

Из теорем 2.56 и 2.1 тотчас вытекает следующее утверждение.

Т е о р е м а 2.57. *Функция $f: D(f) \to \mathbf{R}$ выпукла на $D(f)$ тогда и только тогда, когда для любого натурального числа m выполняются соотношения*

$$\left.\begin{array}{l} x^1, \ldots, x^m \in D(f) \\ (\lambda_1, \ldots, \lambda_m)' \in P_+^m \end{array}\right\} \Rightarrow f\left(\sum_{i=1}^{m} \lambda_i x^i\right) \leqslant \sum_{i=1}^{m} \lambda_i f(x^i). \tag{2.63}$$

Установленную выше связь между f и $\operatorname{epi} f$ мы используем для доказательства следующей теоремы.

Т е о р е м а 2.58. *Пусть $(f_i)_{i \in Z}$, $f_i: D(f) \to \mathbf{R}$ $\forall i \in Z$, – произвольное семейство выпуклых функций. Тогда функция $f: D(f) \to \mathbf{R}$ вида $f(x) = \sup\limits_{i \in Z} f_i(x)$ выпукла на $D(f)$.*

Д о к а з а т е л ь с т в о. Для произвольного $i \in Z$ f_i – выпуклая функция, а $\operatorname{epi} f_i$ – выпуклое множество. Поэтому в силу теоремы 2.3 множество $\operatorname{epi} f = \bigcap\limits_{i \in Z} \operatorname{epi} f_i$ выпукло и, следовательно, функция f тоже выпукла.

О выпуклости сложной функции можно высказать следующее утверждение.

Т е о р е м а 2.59. *Пусть $f: D(f) \to \mathbf{R}$ выпукла на $D(f) \subseteq \mathbf{R}^n$, $g: D(g) \to \mathbf{R}$ монотонно возрастает и выпукла на $D(g) \subseteq \mathbf{R}$ и $f(D(f)) \subseteq D(g)$. Тогда функция $F: D(f) \to \mathbf{R}$ вида $F(x) = g(f(x))$ выпукла на $D(f)$.*

Д о к а з а т е л ь с т в о. Для произвольных $x^1, x^2 \in D(f)$ и $(\lambda_1, \lambda_2)' \in P_+^2$ в силу выпуклости f и монотонности g имеем

$$g(f(\lambda_1 x^1 + \lambda_2 x^2)) \leqslant g(\lambda_1 f(x^1) + \lambda_2 f(x^2)),$$

и так как g выпукла, то

$$g(f(\lambda_1 x^1 + \lambda_2 x^2)) \leqslant \lambda_1 g(f(x^1)) + \lambda_2 g(f(x^2)). \tag{2.64}$$

Получим следствие теоремы 2.59.

Т е о р е м а 2.60. *Пусть $D(f) \subseteq \mathbf{R}^n$ – выпуклое множество. Если функция f выпукла на $D(f)$ и $f(x) \geqslant 0$ для всех $x \in D(f)$, то и функция f^α при $\alpha \geqslant 1$ выпукла на $D(f)$.*

Легко усматривается также справедливость следующего утверждения.

Теорема 2.61. *Пусть* $g: D(g) \to \mathbf{R}$, $D(g) \subseteq \mathbf{R}^m$, *– выпуклая функция и* $A: \mathbf{R}^n \to \mathbf{R}^m$ *– линейное преобразование, для которого* $D(f) = \{x: Ax \in D(g)\} \neq \phi$. *Тогда функция* $f: \mathbf{R}^n \to \mathbf{R}$ *вида* $f(x) = g(Ax)$ *выпукла на* $D(f)$.

2.2.2. О минимуме выпуклых функций. Для выпуклых функций имеет место следующая основная теорема.

Теорема 2.62. *Пусть* $f: D(f) \to \mathbf{R}$ *выпукла на* $D(f) \subseteq \mathbf{R}^n$. *Тогда любая точка локального минимума функции* f *на* $D(f)$ *есть в то же время и точка глобального минимума* f *на* $D(f)$.

Доказательство. Проведем доказательство от противного. Допустим, что f достигает локального минимума в точке x^1, а глобального в x^0, и при этом $f(x^1) > f(x^0)$. Тогда для произвольного вектора $(\lambda_1, \lambda_2)' \in \mathrm{ri}\, P_+^2$ имеем

$$f(\lambda_1 x^0 + \lambda_2 x^1) \leqslant \lambda_1 f(x^0) + \lambda_2 f(x^1) < \lambda_1 f(x^1) + \lambda_2 f(x^1) = f(x^1).$$

Если $\lambda_1 \to +0$, то $\lambda_1 x^0 + \lambda_2 x^1 \to x^1$. Поэтому в любой окрестности x^1 найдутся точки $x \in D(f)$, для которых $f(x) < f(x^1)$. Но это противоречит предположению, согласно которому f достигает в x^1 локального минимума.

Из теоремы 2.62 получаем следующее утверждение.

Теорема 2.63. *Пусть функция* $f: D(f) \to \mathbf{R}$ *выпукла на* $D(f) \subseteq \mathbf{R}^n$. *Тогда множество* D^0 *всех точек, в которых* f *достигает на* $D(f)$ *локального (а тем самым, и глобального) минимума, выпукло.*

Существование глобального минимума выпуклой функции f обсуждается в следующей теореме.

Теорема 2.64. *Пусть множество* $D(f) \subseteq \mathbf{R}^n$ *компактно и функция* $f: D(f) \to \mathbf{R}$ *непрерывна и строго выпукла на* $D(f)$. *Тогда* f *достигает на* $D(f)$ *(глобального) минимума в единственной точке* x^0.

Доказательство. Так как $D(f)$ компактно, а f непрерывна, то f достигает минимума на $D(f)$. Допустим, что f достигает этого минимума в точках $x^0, x^1 \in D(f)$ и $x^0 \neq x^1$. Тогда для произвольного вектора $(\lambda_1, \lambda_2)' \in \mathrm{ri}\, P_+^2$ имеем

$$f(\lambda_1 x^0 + \lambda_2 x^1) < \lambda_1 f(x^0) + \lambda_2 f(x^1) = \lambda_1 f(x^0) + \lambda_2 f(x^0) = f(x^0).$$

Это противоречит предположению о том, что f достигает минимума в x^0.

Формулировка теоремы 2.64 приводит к мысли о том, что выпуклая функция не обязательно непрерывна. Мы исследуем этот вопрос более подробно в следующем пункте.

2.2.3. Непрерывность выпуклых функций. Как показывают простые примеры, не каждая выпуклая на $D(f) \subseteq \mathbf{R}^n$ функция является в то же время непрерывной на $D(f)$.

Пример 2.16. Функция

$$f(x) = \begin{cases} 0, & \text{если } 0 \leqslant x < 1, \\ 1, & \text{если } x = 1, \end{cases}$$

выпукла на $[0, 1]$, но не непрерывна.

Однако справедливо следующее утверждение.

Теорема 2.65. *Пусть* $D(f) \subseteq \mathbf{R}^n$ *– выпуклое множество и* $\operatorname{int} D(f) \neq \phi$. *Если* f*– выпуклая на* $D(f)$ *функция, то она непрерывна на* $\operatorname{int} D(f)$.

Доказательство. Пусть произвольно выбраны точка $x^0 \in \operatorname{int} D(f)$ и число $\epsilon > 0$. Рассмотрим окрестность $S(x^0, \rho) \subseteq D(f)$ точки x^0. Обозначим через $e^1, \ldots, e^n$ единичные базисные векторы пространства $\mathbf{R}^n$. Если $\delta \in [0, \rho)$, то, принимая во внимание разложение $x^0 + \delta e^j = (1 - \delta/\rho)\, x^0 + (\delta/\rho)\ (x^0 + \rho e^j)$, в силу выпуклости f получим для $j = 1, \ldots, n$

$$f(x^0 + \delta e^j) \leqslant \left(1 - \frac{\delta}{\rho}\right) f(x^0) + \frac{\delta}{\rho}\, f(x^0 + \rho e^j),$$

откуда вытекает

$$f(x^0 + \delta e^j) - f(x^0) \leqslant \frac{\delta}{\rho}\, [f(x^0 + \rho e^j) - f(x^0)]. \tag{2.65}$$

Аналогично для $j = 1, \ldots, n$ выполняется неравенство

$$f(x^0 - \delta e^j) - f(x^0) \leqslant \frac{\delta}{\rho}\, [f(x^0 - \rho e^j) - f(x^0)]. \tag{2.66}$$

Из (2.65) и (2.66) следует существование числа $\delta_1 \in (0, \rho)$ такого, что для $j = 1, \ldots, n$ имеем

$$f(x^0 + \delta e^j) - f(x^0) < \epsilon \qquad \forall \delta,\ |\delta| < \delta_1. \tag{2.67}$$

Для произвольного вектора $h = (h_1, \ldots, h_n)' \in \mathbf{R}^n$ в силу теоремы 2.57 получим

$$f(x^0 + h) - f(x^0) = f\left(x^0 + \frac{1}{n} \sum_{j=1}^{n} n h_j e^j\right) - f(x^0) =$$

$$= f\left(\sum_{j=1}^{n} \frac{1}{n}\ [x^0 + n h_j e^j]\right) - f(x^0) \leqslant \sum_{j=1}^{n} \frac{1}{n}\ f(x^0 + n h_j e^j) -$$

$$- f(x^0) = \frac{1}{n} \sum_{j=1}^{n} f(x^0 + n h_j e^j) - f(x^0).$$

Положим $\delta_0 = \delta_1 / n > 0$. Для $\| h \| < \delta_0$ имеем $| n h_j | < \delta_1$, $j = 1, \ldots, n$. Отсюда следует

$$f(x^0 + h) - f(x^0) < \epsilon \qquad \forall h \in \mathbf{R}^n, \quad \| h \| < \delta_0. \tag{2.67}$$

Далее, справедливо соотношение

$$f(x^0) = f\left(\frac{1}{2}\ (x^0 + h) + \frac{1}{2}\ (x^0 - h)\right) \leqslant \frac{1}{2}\ f(x^0 + h) + \frac{1}{2}\ f(x^0 - h),$$

и поэтому $f(x^0) - f(x^0 + h) \leqslant f(x^0 - h) - f(x^0)$. Согласно (2.67) для $\| h \| < \delta_0$ получим

$$f(x^0) - f(x^0 + h) \leqslant f(x^0 - h) - f(x^0) < \epsilon. \tag{2.68}$$

Из (2.67) и (2.68) следует

$$|f(x^0 + h) - f(x^0)| < \epsilon \qquad \forall h \in \mathbf{R}^n, \qquad \| h \| < \delta_0.$$

Таким образом, f непрерывна в x^0.

Простым следствием теоремы 2.65 является утверждение о том, что любая выпуклая функция $f\colon \mathbf{R}^n \to \mathbf{R}$ непрерывна.

С помощью дополнительных рассуждений (см. замечание к теореме 2.22) удается получить следующее обобщение теоремы 2.65.

Теорема 2.66. *Пусть $D(f) \subseteq \mathbf{R}^n$ – выпуклое множество. Если функция f выпукла на $D(f)$, то она непрерывна на множестве* $\operatorname{ri} D(f)$.

Теперь мы введем понятие полунепрерывной функции. Функция f: $D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$, называется *полунепрерывной снизу в точке* $x^0 \in D(f)$, если для любого $\epsilon > 0$ существует $\delta = \delta(\epsilon;\ x^0)$ такое, что

$$f(x^0) - \epsilon \leqslant f(x) \qquad \forall x \in D(f) \cap S(x^0; \delta). \tag{2.69}$$

Это условие эквивалентно тому, что

$$f(x^0) \leqslant \varliminf_{\substack{x \to x^0 \\ x \in D(f)}} f(x) = \lim_{\delta \to +0} \inf_{x \in S(x^0, \delta) \cap D(f)} f(x). \tag{2.70}$$

По определению нижнего предела для любого $x^0 \in D(f)$ выполняется неравенство

$$f(x^0) \geqslant \varliminf_{\substack{x \to x^0 \\ x \in D(f)}} f(x),$$

так что (2.70) может быть записано и в виде

$$f(x^0) = \varliminf_{\substack{x \to x^0 \\ x \in D(f)}} f(x). \tag{2.71}$$

Функция f называется *полунепрерывной снизу на множестве* $D(f)$, если условие (2.70) (или (2.71)) выполняется для любой точки $x^0 \in D(f)$.

Функция f: $D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$, называется *полунепрерывной сверху в точке* $x^0 \in D(f)$ (соответственно *на множестве* $D(f)$), если $-f$ полунепрерывна снизу в x^0 (соответственно на $D(f)$).

Если функция f в точке $x^0 \in D(f)$ полунепрерывна сверху и снизу, то она непрерывна в x^0, потому что тогда

$$\varliminf_{\substack{x \to x^0 \\ x \in D(f)}} f(x) = \varlimsup_{\substack{x \to x^0 \\ x \in D(f)}} f(x) = f(x^0).$$

Приведенная в примере 2.16 функция не является полунепрерывной снизу в точке $x^0 = 1$, но полунепрерывна сверху в этой точке.

Тот факт, что полунепрерывная снизу выпуклая функция не обязательно непрерывна, показывает следующий пример.

Пример 2.17. Для функции f: $D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^2$, вида

$$f(x) = \begin{cases} \dfrac{x_1^2 + x_2^2}{2x_1}, & \text{если } x_1 > 0, \\ 0, & \text{если } x_1 = 0,\ x_2 = 0, \end{cases}$$

на рис. 2.17 изображено несколько линий уровня. С одной стороны, имеем

$$\varliminf_{\substack{x \to 0 \\ x \in D(f)}} f(x) = 0 = f(0),$$

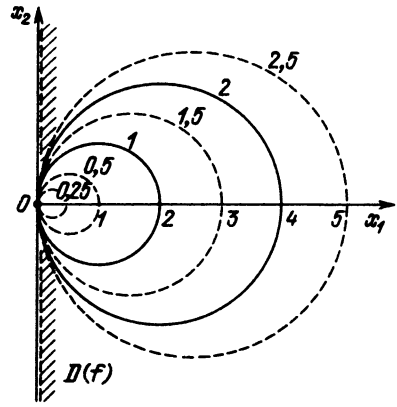


Рис. 2.17.

с другой, для $x_1 = x_2^2 > 0$ выполняется

$$\lim_{x_2 \to 0} \frac{x_2^4 + x_2^2}{2x_2^2} = \frac{1}{2}.$$

Отсюда

$$\overline{\lim_{\substack{x \to 0 \\ x \in D(f)}}} f(x) > \underline{\lim_{\substack{x \to 0 \\ x \in D(f)}}} f(x),$$

и, таким образом, заданная функция не является непрерывной в $x^0 = 0$. В действительности в любой окрестности начала координат она может принимать сколь угодно большие значения.

Полунепрерывность функции f снизу имеет важное значение при исследовании замкнутости множества epi f в 2.2.6.

2.2.4. Дифференцируемость выпуклых функций. Вначале исследуем свойства разностных отношений для выпуклых функций.

Теорема 2.67. *Пусть* $f\colon D(f) \to \mathbf{R}, D(f) \subseteq \mathbf{R}^n$, – *выпуклая функция,* $x^0 \in D(f), x^0 - \nu h \in D(f), x^0 + \rho h \in D(f)$ *для некоторых* $\nu > 0, \rho > 0$. *Тогда*

$$\frac{f(x^0) - f(x^0 - \nu h)}{\nu} \leqslant \frac{f(x^0) - f(x^0 - \mu h)}{\mu} \leqslant \frac{f(x^0 + \lambda h) - f(x^0)}{\lambda} \leqslant$$

$$\leqslant \frac{f(x^0 + \rho h) - f(x^0)}{\rho} \qquad \forall \mu \in (0, \nu], \quad \forall \lambda \in (0, \rho]. \tag{2.72}$$

Доказательство. Неравенства (2.72) получаются преобразованием соотношений

$$f(x^0 - \mu h) \leqslant \frac{\nu - \mu}{\nu} f(x^0) + \frac{\mu}{\nu} f(x^0 - \nu h),$$

$$f(x^0) \leqslant \frac{\lambda}{\lambda + \mu} f(x^0 - \mu h) + \frac{\mu}{\lambda + \mu} f(x^0 + \lambda h),$$

$$f(x^0 + \lambda h) \leqslant \frac{\rho - \lambda}{\rho} f(x^0) + \frac{\lambda}{\rho} f(x^0 + \rho h).$$

Из теоремы 2.67 видно, что разностное отношение для выпуклых функций в точке x^0 в направлении h является монотонно возрастающей функцией от λ. Кроме того, в точке x^0 всегда существует в виде собственного или несобственного предела (односторонняя) производная функции f по направлению h:

$$f^+(x^0; h) = \lim_{\lambda \to +0} \frac{f(x^0 + \lambda h) - f(x^0)}{\lambda}.$$

Будем считать, что $f^+(x^0; 0) = 0$.

При выполнении условий теоремы 2.67 справедливо соотношение

$$\frac{f(x^0)-f(x^0-\nu h)}{\nu} \leqslant -f^+(x^0;-h) \leqslant f^+(x^0;h) \leqslant \frac{f(x^0+\rho h)-f(x^0)}{\rho} \tag{2.73}$$

Это утверждение может быть усилено следующим образом.

Теорема 2.68. *Пусть* $f\colon D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$, *– выпуклая функция,* $x^0 \in \operatorname{ri} D(f)$, $x^0 + h \in V(D(f))$. *Тогда существует конечная производная* $f^+(x^0;h)$.

Доказательство. В силу $x^0 + h \in V(D(f))$ существует отрезок прямой $\{y\colon y = x^0 + \lambda h, \lambda \in \mathbf{R}\}$, полностью лежащий в $D(f)$ и содержащий x^0 в качестве относительно внутренней точки. Тогда найдутся числа $\nu > 0$ и $\rho > 0$ такие, что

$$x^0 - \mu h \in D(f), \quad x^0 + \lambda h \in D(f) \quad \forall \mu \in (0, \nu] \quad \forall \lambda \in (0, \rho].$$

Отсюда и из (2.73) вытекает требуемое утверждение.

В частности, определенная на открытом множестве $D(f) \subseteq \mathbf{R}^n$ выпуклая функция $f\colon D(f) \to \mathbf{R}$ в каждой точке $D(f)$ имеет конечную производную по любому направлению.

Справедливо и обратное утверждение.

Теорема. 2.69. *Пусть* $D(f) \subseteq \mathbf{R}^n$, $\operatorname{int} D(f) \neq \phi$ *и* $f\colon D(f) \to \mathbf{R}$ *– выпуклая функция. Производная* $f^+(x^0;h)$ *тогда и только тогда конечна по любому направлению* $h \in \mathbf{R}^n$, *когда* $x^0 \in \operatorname{int} D(f)$.

Производная $f^+(x^0;h)$ является положительно однородной функцией (степени 1) переменной h, потому что для $\alpha > 0$ выполняются равенства

$$f^+(x^0;\alpha h) = \lim_{\lambda \to +0} \frac{f(x^0+\lambda\alpha h)-f(x^0)}{\lambda} =$$

$$= \lim_{\lambda \to +0} \alpha \left(\frac{f(x^0+\lambda\alpha h)-f(x^0)}{\alpha\lambda} \right) = \alpha f^+(x^0;h). \tag{2.74}$$

Более того, имеет место следующее утверждение.

Теорема 2.70. *Пусть* $f\colon D(f) \to \mathbf{R}, D(f) \subseteq \mathbf{R}^n$, *– выпуклая функция,* $x^0 \in \operatorname{ri} D(f)$, $x^0 + h \in V(D(f))$, $x^0 + k \in V(D(f))$. *Тогда*

$$f^+(x^0;h+k) \leqslant f^+(x^0;h) + f^+(x^0;k),$$

т.е. производная по направлению субаддитивна относительно направления.

Доказательство. Для достаточно малого $\lambda > 0$ имеем

$$f(x^0 + \lambda(h+k)) = f\left(\frac{1}{2}(x^0 + 2\lambda h) + \frac{1}{2}(x^0 + 2\lambda k)\right) \leqslant$$

$$\leqslant \frac{1}{2} f(x^0 + 2\lambda h) + \frac{1}{2} f(x^0 + 2\lambda k),$$

поэтому

$$\frac{f(x^0+\lambda(h+k))-f(x^0)}{\lambda} \leqslant \frac{1}{2} \frac{f(x^0+2\lambda h)-f(x^0)}{\lambda} +$$

$$+ \frac{1}{2} \frac{f(x^0+2\lambda k)-f(x^0)}{\lambda},$$

откуда

$$f^+(x^0; h+k) \leqslant \frac{1}{2} f^+(x^0; 2h) + \frac{1}{2} f^+(x^0; 2k) =$$

$$= f^+(x^0; h) + f^+(x^0; k).$$

В силу (2.74) из теоремы 2.70 следует, что для любой выпуклой функции f: $D(f) \to \mathbf{R}, D(f) \subseteq \mathbf{R}^n$, производная по направлению $f^+(x^0; h)$ в любой фиксированной точке $x^0 \in \operatorname{ri} D(f)$ является выпуклой функцией от h на подпространстве $L = V(D(f)) - x^0$.

Обратимся теперь к дифференцируемым выпуклым функциям. Известно, что не всякая функция f: $D(f) \to \mathbf{R}$, имеющая в точке $x^0 \in D(f)$ частные производные по всем переменным, дифференцируема в x^0. Однако для выпуклых функций такое утверждение справедливо, что показывает следующая теорема.

Теорема 2.71. *Пусть f: $D(f) \to \mathbf{R}, D(f) \subseteq \mathbf{R}^n$, – выпуклая функция. Если f имеет в точке $x^0 \in D(f)$ частные производные по всем переменным, то она дифференцируема в x^0.*

Доказательство. Рассмотрим точки $h = (h_1, \ldots, h_n)' \in \mathbf{R}^n$, для которых $x^0 + h \in D(f)$, и положим

$$g(h) = f(x^0 + h) - f(x^0) - \sum_{j=1}^{n} \frac{\partial f(x^0)}{\partial x_j} h_j.$$

Требуется показать, что

$$\lim_{h \to 0} \frac{g(h)}{\|h\|} = 0. \tag{2.75}$$

Очевидно, что функция g выпуклая на множестве $\{h: x^0 + h \in D(f)\}$ и $g(0) = 0$. Из соотношений

$$0 = g(0) = g\left(\frac{1}{2} h + \frac{1}{2}(-h)\right) \leqslant \frac{1}{2} g(h) + \frac{1}{2} g(-h)$$

следует, что

$$g(h) \geqslant -g(-h). \tag{2.76}$$

Положим теперь

$$\epsilon_j(h) = \frac{g(nh_j e^j)}{nh_j}, \qquad j = 1, \ldots, n. \tag{2.77}$$

Имеем

$$\epsilon_j(h) = \frac{f(x^0 + nh_j e^j) - f(x^0)}{nh_j} - \frac{\partial f(x^0)}{\partial x_j}$$

и, в силу выпуклости $g(h)$,

$$g(h) = g\left(\frac{1}{n} \sum_{j=1}^{n} nh_j e^j\right) \leqslant \frac{1}{n} \sum_{j=1}^{n} g(nh_j e^j) = \sum_{j=1}^{n} h_j \epsilon_j(h).$$

Отсюда следует

$$g(h) \leqslant \gamma \Big(\sum_{j=1}^{n} |\epsilon_j(h)|^2 \Big)^{1/2} \| h \| \quad \forall\, h \in \mathbf{R}^n \tag{2.78}$$

и, с учетом (2.76),

$$g(h) \geqslant -g(-h) \geqslant -\Big(\sum_{j=1}^{n} |\epsilon_j(-h)|^2 \Big)^{1/2} \| h \|. \tag{2.79}$$

Так как функция f в точке x^0 имеет частные производные по всем переменным, то $\lim\limits_{h \to 0} \epsilon_j(h) = 0, j = 1, \ldots, n$. Поэтому из (2.78) и (2.79) вытекает (2.75).

Для приложений выпуклых функций оказывается важным следующее утверждение.

Т е о р е м а 2.72. *Пусть $D(f) \subseteq \mathbf{R}^n$ – открытое выпуклое множество и функция f: $D(f) \to \mathbf{R}$ дифференцируема на $D(f)$. Функция f тогда и только тогда выпукла на $D(f)$, когда для любых $x^1, x^2 \in D(f)$, $x^1 \neq x^2$, выполняется одно из условий*:

1) $f(x^2) - f(x^1) \geqslant (x^2 - x^1)' \nabla f(x^1)$, (2.80)

2) $(x^2 - x^1)' [\nabla f(x^2) - \nabla f(x^1)] \geqslant 0$. (2.81)

Д о к а з а т е л ь с т в о. 1. Условие (2.80) необходимо для выпуклости f на $D(f)$. Действительно, так как f дифференцируема, то

$$h' \nabla f(x^1) = f^+(x^1; h) = -f^+(x^1; -h).$$

Из теоремы 2.67 тогда следует выполнение соотношения

$$\frac{f(x^1 + \rho h) - f(x^1)}{\rho} \geqslant f^+(x^1, h) = h' \nabla f(x^1)$$

для любого $\rho > 0$ такого, что $x^1 + \rho h \in D(f)$. Отсюда

$$f(x^1 + \rho h) \geqslant f(x^1) + \rho h' \nabla f(x^1).$$

Полагая $x^2 = x^1 + \rho h$, получим (2.80).

2. Условие (2.80) достаточно для выпуклости f на $D(f)$. В самом деле, построив для произвольного вектора $(\lambda_1, \lambda_2)' \in P_+^2$ точку $x^3 = \lambda_1 x^1 + \lambda_2 x^2$, получим

$$f(x^1) - f(x^3) \geqslant (x^1 - x^3)' \nabla f(x^3), \quad f(x^2) - f(x^3) \geqslant (x^2 - x^3)' \nabla f(x^3),$$

и поэтому

$$\lambda_1 f(x^1) + \lambda_2 f(x^2) \geqslant f(x^3) + [\lambda_1 (x^1 - x^3) + \lambda_2 (x^2 - x^3)]' \nabla f(x^3) =$$

$$= f(x^3) + [\lambda_1 x^1 + \lambda_2 x^2 - x^3]' \nabla f(x^3) = f(x^3) = f(\lambda_1 x^1 + \lambda_2 x^2)$$

$\forall\, x^1, x^2 \in D(f)$.

3. Утверждение теоремы, относящееся к условию (2.81), доказывается так же легко.

Неравенство (2.80) наглядно отражает тот факт, что надграфик дифференцируемой выпуклой функции f лежит всегда "выше" касательной гиперплоскости к графику f в произвольной точке.

Утверждения, аналогичные теореме 2.72, имеют место и для строго выпуклых функций.

Теорема 2.73. *Пусть $D(f) \subseteq \mathbf{R}^n$ – открытое выпуклое множество и функция f: $D(f) \to \mathbf{R}$ дифференцируема на $D(f)$. Функция f тогда и только тогда строго выпукла на $D(f)$, когда для любых $x^1, x^2 \in D(f)$, $x^1 \neq x^2$, выполняется одно из условий:*

$$1)\ f(x^2) - f(x^1) > (x^2 - x^1)'\nabla f(x^1); \tag{2.82}$$

$$2)\ (x^2 - x^1)'[\nabla f(x^2) - \nabla f(x^1)] > 0. \tag{2.83}$$

Для дважды непрерывно дифференцируемой функции f: $D(f) \to \mathbf{R}$, определенной на открытом множестве $D(f) \subseteq \mathbf{R}^n$, гессиан

$$\nabla^2 f(x) = \left[\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right]_{i,j=1,\ldots,n} \tag{2.84}$$

является симметрической матрицей, и для любого фиксированного $x^0 \in D(f)$ имеет место разложение

$$f(x) = f(x^0) + (x - x^0)'\nabla f(x^0) + \frac{1}{2}(x - x^0)'\nabla^2 f(x^0)(x - x^0) +$$

$$+ \| x - x^0 \|^2 \theta(x^0, x) \qquad \forall x \in D(f). \tag{2.85}$$

где $\theta(x^0, x) \to 0$ при $x \to x^0$.

Если вместе с точками x^1, x^2 в $D(f)$ полностью лежит сегмент $[x^1, x^2]$, то по теореме Тейлора для некоторого $\lambda \in (0, 1)$ имеем

$$f(x^2) = f(x^1) + (x^2 - x^1)'\nabla f(x^1) + \frac{1}{2}(x^2 - x^1)'\nabla^2 f(x^1 + \lambda[x^2 - x^1])(x^2 - x^1). \tag{2.86}$$

Дважды непрерывно дифференцируемые выпуклые функции характеризуются тем, что их матрица Гессе положительно полуопределена.

Теорема 2.74. *Пусть $D(f) \subseteq \mathbf{R}^n$ – открытое выпуклое множество и функция f: $D(f) \to \mathbf{R}$ дважды непрерывно дифференцируема на $D(f)$. Тогда справедливо утверждение.*

f выпукла на $D(f) \Longleftrightarrow$ матрица $\nabla^2 f(x)$ положительно полуопределена для всех $x \in D(f)$. (2.87)

Доказательство. 1. Пусть матрица $\nabla^2 f(x)$ положительно полуопределена. Тогда в силу (2.86) для произвольных $x^1, x^2 \in D(f)$ и некоторого $\lambda \in [0, 1]$ имеем

$$f(x^2) - f(x^1) - (x^2 - x^1)'\nabla f(x^1) =$$

$$= \frac{1}{2}(x^2 - x^1)'\nabla^2 f(x^1 + \lambda[x^2 - x^1])(x^2 - x^1) \geqslant 0.$$

По теореме 2.72 функция f выпукла на $D(f)$.

2. Пусть f выпукла на $D(f)$ и пусть заданы произвольные точки $x^1 \in D(f)$ и $h \in \mathbf{R}^n$. Так как множество $D(f)$ открыто, то найдется число $\theta_0 > 0$ такое, что $x^1 + \theta h \in D(f)$ $\forall \theta \in [0, \theta_0]$. Поэтому, заменив в (2.80) x^2 на $x^1 + \theta h$, будем иметь

$$f(x^1 + \theta h) - f(x^1) - \theta h'\nabla f(x^1) \geqslant 0 \qquad \forall \theta \in [0, \theta_0].$$

Учитывая (2.86), получим

$$h'\nabla^2 f(x^1 + \lambda\theta h)h \geqslant 0 \qquad \forall\, \theta \in [0, \theta_0],$$

где $\lambda \in (0, 1)$. Отсюда и из условий теоремы при $\theta \to +0$ вытекает $h'\nabla^2 f(x^1)h \geqslant 0$. Так как $x^1 \in D(f)$ и $h \in \mathbf{R}^n$ были выбраны произвольно, то матрица $\nabla^2 f(x)$ положительно полуопределена для любого $x \in D(f)$.

Если в условиях теоремы 2.74 матрица Гессе функции f положительно определена для любого $x \in D(f)$, то f строго выпукла на $D(f)$. Обратное же утверждение места не имеет.

Пример 2.18. Функция

$$f(x) = x^4, \qquad x \in D(f) = \mathbf{R},$$

строго выпукла, но $\nabla^2 f(0) = 0$.

Пример 2.19. С помощью теоремы 2.74 легко проверяется выпуклость следующих функций f: $D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$:

а) $f(x) = x^\alpha$, $\quad D(f) = \mathbf{R}_+$, $\quad \alpha \in [1, \infty)$;

б) $f(x) = -x^\alpha$, $\quad D(f) = \mathbf{R}_+$, $\quad \alpha \in [0, 1]$;

в) $f(x) = e^{\alpha x}$, $\quad D(f) = \mathbf{R}$, $\quad \alpha \in \mathbf{R}$;

г) $f(x) = x^\alpha$, $\quad D(f) = \operatorname{int} \mathbf{R}_+$, $\quad \alpha \in (-\infty, 0]$;

д) $f(x) = -\ln x$, $\quad D(f) = \operatorname{int} \mathbf{R}_+$.

Вернемся еще раз к вопросу о минимуме выпуклых функций (см. п. 2.2.2).

Если множество $D(f)$ открыто, то, как известно, условие $\nabla f(x^0) = 0$, $x^0 \in D(f)$, вообще говоря, лишь необходимо для существования локального экстремума функции f на $D(f)$. В случае же выпуклых функций справедливо следующее утверждение.

Теорема 2.75. *Пусть $D(f) \subseteq \mathbf{R}^n$ – открытое выпуклое множество и f: $D(f) \to \mathbf{R}$ – выпуклая дифференцируемая на $D(f)$ функция. Если $x^0 \in D(f)$ и $\nabla f(x^0) = 0$, то f достигает в точке x^0 глобального минимума на $D(f)$.*

Доказательство. По теореме 2.72 имеем

$$f(x) - f(x^0) \geqslant (x - x^0)'\nabla f(x^0) = 0 \qquad \forall\, x \in D(f).$$

2.2.5. Субдифференциал выпуклых функций. Этот раздел содержит утверждения о гиперплоскостях, опорных к множеству epi f для выпуклых (не обязательно дифференцируемых) функций.

Множество

$$H = \left\{ \begin{pmatrix} x \\ \xi \end{pmatrix} \in \mathbf{R}^{n+1} : \; a'x - \beta\xi = \alpha \right\},$$

$a \in \mathbf{R}^n$; $\alpha, \beta \in \mathbf{R}$, представляет собой гиперплоскость в пространстве $\mathbf{R}^{n+1}$. Если $\beta = 0$, то гиперплоскость называется *вертикальной*, в противном случае – *невертикальной*.

Без ограничения общности любая невертикальная гиперплоскость пространства $\mathbf{R}^{n+1}$ может быть представлена*) в виде

$$H(a, \alpha) = \left\{ \begin{pmatrix} x \\ \xi \end{pmatrix} \in \mathbf{R}^{n+1} : \; a'x - \xi = \alpha \right\}. \tag{2.88}$$

Она пересекает ось ξ в точке $\begin{pmatrix} 0 \\ -\alpha \end{pmatrix}$. Направление оси ξ при этом называется (по аналогии со случаем $n = 1$) вертикальным направлением, а сама ось ξ – вертикальной осью.

Гиперплоскость $H(a, \alpha)$ определяет полупространства

$$H_+(a, \alpha) = \left\{ \begin{pmatrix} x \\ \xi \end{pmatrix} \in \mathbf{R}^{n+1} : \; a'x - \xi \geqslant \alpha \right\},$$

$$H_-(a, \alpha) = \left\{ \begin{pmatrix} x \\ \xi \end{pmatrix} \in \mathbf{R}^{n+1} : \; a'x - \xi \leqslant \alpha \right\}. \tag{2.89}$$

Гиперплоскость $H(a, \alpha)$ тогда и только тогда является невертикальной опорной гиперплоскостью к множеству $\operatorname{epi} f$ в опорной точке $(x^0, f(x^0))'$, когда

$$H(a, \alpha) = \left\{ \begin{pmatrix} x \\ \xi \end{pmatrix} \in \mathbf{R}^{n+1} : \; a'x - \xi = a'x^0 - f(x^0) \right\}$$

и, одновременно,

$$H_-(a, \alpha) \supseteq \operatorname{epi} f,$$

т.е. $a'x - \xi \leqslant a'x^0 - f(x^0)$ для всех $\begin{pmatrix} x \\ \xi \end{pmatrix}$ таких, что $\xi \geqslant f(x)$, $x \in D(f)$. В частности,

$$a'x - f(x) \leqslant a'x^0 - f(x^0) \quad \forall x \in D(f).$$

Таким образом, множество невертикальных опорных гиперплоскостей к $\operatorname{epi} f$ в опорной точке $(x^0, f(x^0))'$ характеризуется множеством

$$\partial f(x^0) = \{ a \in \mathbf{R}^n : \; f(x) \geqslant f(x^0) + a'(x - x^0) \;\; \forall x \in D(f) \}. \tag{2.90}$$

Назовем множество $\partial f(x^0)$ *субдифференциалом* функции f в точке x^0, а элементы $\partial f(x^0)$ – *субградиентами* f в точке x^0 (рис. 2.18).

Т е о р е м а 2.76. *Пусть* f: $D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$, – *произвольная функция,* $x^0 \in D(f)$ *и* $\partial f(x^0) \neq \phi$. *Тогда множество* $\partial f(x^0)$ *выпукло и замкнуто.*

Д о к а з а т е л ь с т в о. 1. Пусть $y^1, y^2 \in \partial f(x^0)$. Тогда

$$f(x) \geqslant f(x^0) + y^{1\prime}(x - x^0) \quad \forall x \in D(f),$$

$$f(x) \geqslant f(x^0) + y^{2\prime}(x - x^0) \quad \forall x \in D(f).$$

---

*) Сравните с определением гиперплоскости $E(a, \alpha) \subseteq \mathbf{R}^n$, введенным в примере 2.1. Следуя принятому ранее способу обозначения, мы должны были бы для гиперплоскости (2.88) написать $E\left(\begin{pmatrix} a \\ -1 \end{pmatrix}, \alpha\right)$. Однако простоты ради мы будем применять для невертикальных гиперплоскостей пространства $\mathbf{R}^{n+1}$ запись $H(a, \alpha)$.

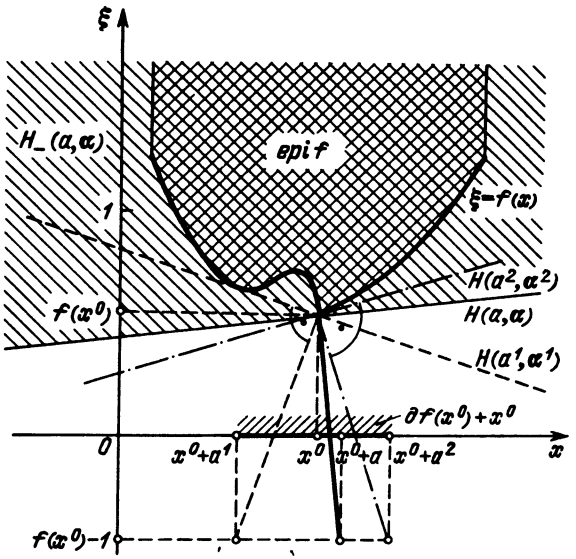


Рис. 2.18.

Поэтому для произвольного вектора $(\lambda_1, \lambda_2)' \in P_+^2$ имеем

$$f(x) \geqslant f(x^0) + (\lambda_1 y^1 + \lambda_2 y^2)'(x - x^0) \quad \forall x \in D(f).$$

2) Пусть $\{y^i\} \subset \partial f(x^0)$ и $y^i \to y^0$. В силу

$$f(x) \geqslant f(x^0) + y^{i\prime}(x - x^0) \quad \forall x \in D(f), \quad i = 1, 2, \ldots,$$

имеем также

$$f(x) \geqslant f(x^0) + y^{0\prime}(x - x^0) \quad \forall x \in D(f),$$

и отсюда $y^0 \in \partial f(x^0)$.

Следующая теорема посвящена вопросу существования субградиентов в точке $x^0 \in D(f)$ и тем самым – вопросу существования невертикальных опорных гиперплоскостей в точке $(x^0, f(x^0))'$.

Т е о р е м а 2.77. *Пусть* f: $D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$, – *выпуклая на* $D(f)$ *функция. В каждой точке* $x^0 \in \operatorname{ri} D(f)$ *существует субградиент, т.е.* $\partial f(x^0) \neq \phi$.

Д о к а з а т е л ь с т в о. Применим к множеству $C = \operatorname{epi} f - \begin{pmatrix} x^0 \\ f(x^0) \end{pmatrix}$ теорему 2.33.

В случае $x^0 \notin \operatorname{ri} D(f)$ множество $\partial f(x^0)$ может быть пусто.

П р и м е р 2.20. Очевидно, что для выпуклой функции $f(x) = -\sqrt{x}$, $x \in \mathbf{R}_+$, выполняется $\partial f(0) = \phi$.

Следующая теорема указывает некоторые условия, эквивалентные существованию субградиентов. Эти эквивалентности имеют важное значение, в частности, при исследовании устойчивости задач нелинейного программирования (см. 5.4 и 5.9).

Т е о р е м а 2.78. *Пусть* f: $D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$, – *выпуклая функция и* $x^0 \in D(f)$. *Следующие утверждения эквивалентны:*

1) $\partial f(x^0) \neq \phi$;

2) *существует число* $M < 0$ *такое, что*

$$\frac{f(x) - f(x^0)}{\| x - x^0 \|} \geqslant M \quad \forall x \in D(f), \quad x \neq x^0;$$

3) *существует число* $M < 0$ *такое, что*

$$f^+(x^0; h) \geqslant M \quad \forall h \in K(D(f) - x^0), \quad \| h \| = 1,$$

где

$$K(D(f) - x^0) = \{h: \exists \lambda > 0, \ x^0 + \lambda h \in D(f)\};$$

4) *существует число* $M < 0$ *такое, что*

$$f^+(x^0; h) \geqslant M \quad \forall h \in K(D(f) - x^0), \quad 0 < \| h \| \leqslant 1.$$

Д о к а з а т е л ь с т в о. 1) $\Rightarrow$ 2). Имеем

$$y \in \partial f(x^0) \Rightarrow f(x) \geqslant f(x^0) + y'(x - x^0) \quad \forall x \in D(f),$$

$$\Rightarrow f(x) - f(x^0) \geqslant - \| y \| \| x - x^0 \| \quad \forall x \in D(f),$$

и, тем самым,

$$\frac{f(x) - f(x^0)}{\| x - x^0 \|} \geqslant M \quad \forall x \in D(f), \quad x \neq x^0,$$

где $M = \min \{-1, - \| y \|\}$.

2) $\Rightarrow$ 3). Для $x = x^0 + \lambda h \in D(f)$, где $\lambda > 0$ и $\| h \| = 1$, получим из 2)

$$\frac{f(x^0 + \lambda h) - f(x^0)}{\lambda} \geqslant M.$$

В силу (2.72) отсюда следует существование $f^+(x^0; h)$ и выполнение неравенства

$$f^+(x^0; h) \geqslant M \quad \forall h \in K(D(f) - x^0), \quad \| h \| = 1.$$

3) $\Rightarrow$ 4). Для любого $h \in K(D(f) - x^0)$, $0 < \| h \| \leqslant 1$, существуют, очевидно, число $\mu > 0$ и вектор $k \in K(D(f) - x^0)$, $\| k \| = 1$, такие, что $h = \mu k$. Тогда согласно (2.74) имеем

$$f^+(x^0; h) = f^+(x^0; \mu k) = \mu f^+(x^0; k) \geqslant M$$

$$\forall k \in K(D(f) - x^0), \quad 0 < \| h \| \leqslant 1.$$

4) $\Rightarrow$ 2). Это утверждение получается непосредственно из теоремы 2.67.

2) $\Rightarrow$ 1). Множества

$$C_1 = \operatorname{epi} f, \quad C_2 = \left\{ \begin{pmatrix} x \\ \xi \end{pmatrix} \in \mathbf{R}^{n+1}: \ \xi < f(x^0) + M \| x - x^0 \| \right\}$$

выпуклы. Кроме того, C_2 открыто и $C_1 \cap C_2 = \phi$. Следовательно, с учетом замечания в конце теоремы 2.34 существует гиперплоскость

$$H = \left\{ \begin{pmatrix} x \\ \xi \end{pmatrix} \in \mathbf{R}^{n+1}: \ y'x + \eta\xi = \alpha \right\},$$

строго разделяющая C_1 и C_2 :

$$y'x + \eta\xi \leqslant \alpha \quad \forall (x, \xi)' \in C_1,$$

$$y'x + \eta\xi > \alpha \quad \forall (x, \xi)' \in C_2.$$

Из условия $(x^0, f(x^0))' \in C_1$ имеем $y'x^0 + \eta f(x^0) \leqslant \alpha$; в силу $(x^0, \rho + f(x^0))' \in C_2 \ \forall \rho < 0$, получим $y'x^0 + \eta f(x^0) > \alpha - \eta\rho \quad \forall \rho < 0$. Отсюда следует $\eta < 0$, а также $y'x^0 + \eta f(x^0) \geqslant \alpha$ и, тем самым, $y'x^0 + \eta f(x^0) = \alpha$.

Для $y^0 = y/\eta$ выполняется неравенство

$$\xi \geqslant f(x^0) - y^{0\prime}(x - x^0) \quad \forall (x, \xi)' \in C_1,$$

и отсюда $-y^0 \in \partial f(x^0)$.

Исследуем теперь для выпуклых функций связи между субдифференциалом и производными по направлению.

Т е о р е м а 2.79. *Пусть* f: $D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$, *– выпуклая функция. Тогда для выполнения соотношения* $a \in \partial f(x^0)$ *необходимо и достаточно, чтобы* $f^+(x^0; h) \geqslant a'h \ \forall h \in K(D(f) - x^0)$.

Д о к а з а т е л ь с т в о. 1. Пусть $a \in \partial f(x^0)$. Тогда

$$f(x^0 + \lambda h) - f(x^0) \geqslant \lambda a'h \tag{2.91}$$

для всех $\lambda > 0$ таких, что $\lambda h \in D(f) - x^0$, и отсюда

$$f^+(x^0; h) \geqslant a'h \quad \forall h \in K(D(f) - x^0). \tag{2.92}$$

2. Пусть выполнено (2.92). Тогда согласно (2.73) справедливо (2.91) и, тем самым, $a \in \partial f(x^0)$.

В частности, для любого фиксированного $h \in K(D(f) - x^0)$ справедливо

$$f^+(x^0; h) \geqslant \sup_{a \in \partial f(x^0)} a'h. \tag{2.93}$$

П р и м е р 2.21. Пусть задана функция f: $D(f) \to \mathbf{R}$, $D(f) = \{x \in \mathbf{R}^2 : x_1 \geqslant 0\}$, вида $f(x) = -\sqrt{x_1}$. Тогда для $h = (0, 1)'$ имеем, с одной стороны, $f^+(0; h) = 0$, а с другой, $\partial f(0) = \phi$ (см. пример 2.20). Если условимся, что $\sup\{\phi\} = -\infty$, то в соответствии с (2.93) будем иметь

$$0 = f^+(0; h) \geqslant \sup_{a \in \partial f(0)} a'h = -\infty.$$

Выведем теперь достаточное условие для того, чтобы в (2.93) имел место знак равенства. Подробное обсуждение этого вопроса можно найти в [74].

Т е о р е м а 2.80. *Пусть* $D(f) \subseteq \mathbf{R}^n$, $\operatorname{int} D(f) \neq \phi$, *и* f: $D(f) \to \mathbf{R}$ *– выпуклая функция. Тогда для любого* $x^0 \in \operatorname{int} D(f)$ *субдифференциал* $\partial f(x^0)$ *является компактным множеством, и при этом*

$$f^+(x^0; h) = \max_{a \in \partial f(x^0)} a'h. \tag{2.94}$$

Д о к а з а т е л ь с т в о. 1. Докажем вначале компактность $\partial f(x^0)$. В силу того, что $x^0 \in \operatorname{int} D(f)$, производная $f^+(x^0; h)$ конечна для любого $h \in \mathbf{R}^n$. Отсюда с учетом выпуклости функции $f^+(x^0; h)$ по h следует существование числа $M > 0$ такого, что $f^+(x^0; h) \leqslant M$ для всех $h \in \mathbf{R}^n$, у которых $\| h \| = 1$. Поэтому в силу (2.93) $a'h \leqslant M$ для всех $h \in \mathbf{R}^n$ таких,

что $\| h \| = 1$, и всех $a \in \partial f(x^0)$. Следовательно,

$$\| a \| \leqslant M \qquad \forall a \in \partial f(x^0),$$

т.е. $\partial f(x^0)$ ограничено и (в силу теоремы 2.76) замкнуто. Таким образом, $\partial f(x^0)$ компактно.

2. Пусть теперь $C_1 = \operatorname{epi} f$ и для некоторого $k \in K(D(f) - x^0)$

$$C_2 = \left\{ \begin{pmatrix} x \\ \xi \end{pmatrix} : \begin{array}{l} x = x^0 + \lambda h, \quad \lambda \geqslant 0, \\ \xi < f(x^0) + \lambda f^+(x^0; k) \end{array} \right\}.$$

В силу

$$f(x^0 + \lambda k) \geqslant f(x^0) + \lambda f^+(x^0; k) \qquad \forall \lambda \geqslant 0$$

имеем $C_1 \cap C_2 = \phi$. Поэтому согласно теореме 2.34 существует разделяющая C_1 и C_2 гиперплоскость $H = \{(x, \xi)' : a'x + \eta\xi = \alpha\}$, где $(a, \eta)' \neq 0$, и

$$a'x + \eta\xi \geqslant a'(x^0 + \lambda k) + \eta(f(x^0) + \lambda f^+(x^0; k)) \tag{2.95}$$

для всех $(x, \xi)' \in C_1$, $\lambda \geqslant 0$. При $x = x^0$ и $\lambda = 0$ отсюда следует $\eta\xi \geqslant \eta f(x^0)$ и тем самым $\eta \geqslant 0$.

Если положить $\eta = 0$, то в силу (2.95)

$$a'(x^0 - (x^0 + \lambda k)) \geqslant 0 \qquad \forall x \in D(f).$$

При $x^0 \in \operatorname{int} D(f)$ и $\lambda = 0$ отсюда вытекает, что $a'(x - x^0) \geqslant 0$ для любого x из некоторой окрестности точки x^0. Это возможно, лишь если $a = 0$, а тем самым и $(a, \eta)' = 0$, что противоречит существованию разделяющей гиперплоскости. Поэтому мы можем считать, что $\eta > 0$.

При $\xi = f(x)$ и $\lambda = 0$ из (2.95) получим

$$f(x) + \frac{a'}{\eta} x \geqslant f(x^0) + \frac{a'}{\eta} x^0 \qquad \forall x \in D(f), \tag{2.96}$$

и отсюда $-a/\eta \in \partial f(x^0)$. Из (2.95) и (2.96) далее имеем

$$f(x) - f(x^0) \geqslant -\frac{a'}{\eta}(x - x^0) + \lambda\left(f^+(x^0; k) + \frac{a'}{\eta} k\right).$$

При $x = x^0$ и $\lambda > 0$ из последнего неравенства следует

$$f^+(x^0; k) \leqslant -\frac{a'}{\eta} k.$$

В силу

$$f^+(x^0; k) \geqslant \sup_{b \in \partial f(x^0)} b'k$$

окончательно получаем

$$f^+(x^0; h) = \max_{b \in \partial f(x^0)} b'k.$$

Равенство (2.94) для выпуклых функций является обобщением соотношения

$$f^+(x^0; h) = h' \nabla f(x^0)$$

между производной по направлению и градиентом дифференцируемых функций.

Обратным к первому утверждению теоремы 2.80 является следующее высказывание.

Т е о р е м а 2.81. *Пусть* f: $D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$, – *выпуклая функция. Если множество* $\partial f(x^0)$ *компактно, то* $x^0 \in \operatorname{int} D(f)$.

Д о к а з а т е л ь с т в о. В случае $D(f) = \mathbf{R}^n$ утверждение теоремы тривиально. Поэтому пусть $D(f) \subset \mathbf{R}^n$ и $x^0 \in D(f) \backslash \operatorname{int} D(f)$. Тогда в $\mathbf{R}^n$ существует полупространство, которое содержит $D(f)$ и образующая гиперплоскость которого проходит через x^0:

$$a'(x - x^0) \leqslant 0 \quad \forall x \in D(f), \quad a \neq 0.$$

Пусть $b \in \partial f(x^0)$. Тогда

$$f(x) \geqslant f(x^0) + b'(x - x^0) \quad \forall x \in D(f),$$

и, следовательно, для каждого $\lambda \geqslant 0$ выполнятся неравенство

$$f(x) \geqslant f(x^0) + (b + \lambda a)'(x - x^0) \quad \forall x \in D(f).$$

Отсюда следует, что множество $\partial f(x^0)$ не ограничено, а это противоречит условию теоремы.

Для дифференцируемых выпуклых функций может быть указана простая связь между субдифференциалом и градиентом.

Т е о р е м а 2.82. *Пусть* f: $D(f) \to \mathbf{R}$ – *выпуклая функция,* $D(f) \subseteq \mathbf{R}^n$ – *открытое множество и* $x^0 \in D(f)$. *Функция* f *тогда и только тогда дифференцируема в точке* x^0, *когда множество* $\partial f(x^0)$ *состоит из одного элемента.*

Д о к а з а т е л ь с т в о. 1. Пусть функция f дифференцируема в точке x^0. В силу того, что $x^0 \in \operatorname{int} D(f) = D(f)$, по теореме 2.77 существует невертикальная опорная гиперплоскость к $\operatorname{epi} f$ в опорной точке $(x^0, f(x^0))'$:

$$f(x) - f(x^0) \geqslant (x - x^0)'a \quad \forall x \in D(f). \tag{2.97}$$

Так как f дифференцируема в x^0, то, кроме того,

$$f(x) - f(x^0) = (x - x^0)' \nabla f(x^0) + o(\|x - x^0\|) \quad \forall x \in D(f). \tag{2.98}$$

Из (2.97) и (2.98) следует

$$(x - x^0)'(a - \nabla f(x^0)) \leqslant o(\|x - x^0\|) \quad \forall x \in D(f),$$

и отсюда в силу $x^0 \in \operatorname{int} D(f)$ имеем

$$a = \nabla f(x^0).$$

2. Пусть $\partial f(x^0)$ – одноэлементное множество. Тогда в силу (2.94) выполняется соотношение

$$f^+(x^0; h) = \max_{a \in \partial f(x^0)} a'h = a'h \quad \forall h \in \mathbf{R}^n$$

Если положить $h = e^i$, $i = 1, \ldots, n$, то с учетом равенства

$$f^+(x^0; h) = -f^+(x^0; -h) = a'h \qquad \forall h \in \mathbf{R}^n$$

получим существование частных производных от f по всем переменным в точке x^0. Тогда по теореме 2.71 функция f дифференцируема в точке x^0.

Замечания. 1. Если выпуклая функция f дифференцируема в точке $x^0 \in D(f)$, то, как показывает доказательство теоремы 2.82,

$$\partial f(x^0) = \{\nabla f(x^0)\}.$$

2. Если для произвольной (не выпуклой) функции f множество $\partial f(x^0)$ одноэлементное, то отсюда не обязательно вытекает дифференцируемость f в точке x^0, что показывает следующий пример.

Пример 2.22. Для функции

$$g(x) = \begin{cases} \dfrac{\sqrt{x_1^6 x_2^2}}{x_1^4 + x_2^2}, & \text{если } x = (x_1, x_2)' \in \mathbf{R}^2 \setminus (0,0)', \\ 0, & \text{если } x = 0, \end{cases}$$

с одной стороны, имеем

$$\lim_{\substack{\|x\| \to 0 \\ x_2 = x_1^2}} \frac{g(x) - g(0)}{\|x\|} = \frac{1}{2},$$

с другой, $g(x) \geqslant 0 \quad \forall x \in \mathbf{R}^2$, и

$$\lim_{\lambda \to +0} \frac{g(\lambda h)}{\lambda} = g^+(0; h) = 0$$

для произвольного $h \neq 0$. Таким образом, $H(0; 0)$ – единственная опорная гиперплоскость к множеству $\operatorname{epi} g$ в точке $(0, 0, 0)'$.

Теорема 2.83. *Пусть* $f_1 : D(f_1) \to \mathbf{R}$, $D(f_1) \subseteq \mathbf{R}^n$, *и* $f_2 : D(f_2) \to \mathbf{R}$, $D(f_2) \subseteq \mathbf{R}^n$, *– выпуклые функции,* $f = \mu_1 f_1 + \mu_2 f_2$, $\mu_1, \mu_2 \in \mathbf{R}_+$, $D(f) = D(f_1) \cap D(f_2)$. *Если множество* $\operatorname{ri} D(f_1) \cap \operatorname{ri} D(f_2)$ *не пусто, то* $\partial f(x) = \mu_1 \partial f_1(x) + \mu_2 \partial f_2(x)$ *для всех* $x \in D(f)$.

Доказательство. 1. Пусть $x^0 \in D(f)$ – произвольная фиксированная точка и $y = \mu_1 a + \mu_2 b$, где $a \in \partial f_1(x^0)$, $b \in \partial f_2(x^0)$. Тогда

$$\mu_1 f_1(x) + \mu_2 f_2(x) \geqslant \mu_1 f_1(x^0) + \mu_2 f_2(x^0) + (\mu_1 a + \mu_2 b)'(x - x^0)$$

$$\forall x \in D(f_1) \cap D(f_2),$$

т.е. $y \in \partial(\mu_1 f_1(x^0) + \mu_2 f_2(x^0)) = \partial f(x^0)$.

Отсюда следует, что

$$\mu_1 \partial f_1(x) + \mu_2 \partial f_2(x) \subseteq \partial f(x) \qquad \forall x \in D(f).$$

2. Покажем теперь, что при условиях теоремы выполняется соотношение

$$\partial f(x) \subseteq \mu_1 \partial f_1(x) + \mu_2 \partial f_2(x) \qquad \forall x \in D(f).$$

Пусть x^0 – произвольная фиксированная точка из $D(f)$ и $y \in \partial f(x^0)$. Тогда

$$\mu_1 f_1(x) + \mu_2 f_2(x) \geqslant \mu_1 f_1(x^0) + \mu_2 f_2(x^0) + (x - x^0)'y \quad \forall x \in D(f),$$

$$\mu_1 f_1(x) - \mu_1 f_1(x^0) - (x - x^0)'y \geqslant \mu_2 f_2(x^0) - \mu_2 f_2(x) \quad \forall x \in D(f). \quad (2.99)$$

Рассмотрим далее надграфик и подграфик функций, стоящих соответственно в левой и правой частях неравенства (2.99):

$$C_1 = \{(x, \xi)': \ \xi \geqslant \mu_1 f_1(x) - \mu_1 f_1(x^0) - (x - x^0)'y \quad \forall x \in D(f_1)\},$$

$$C_2 = \{(x, \xi)': \ \xi \leqslant \mu_2 f_2(x^0) - \mu_2 f_2(x) \quad \forall x \in D(f_2)\}.$$

В силу (2.99) имеем $\operatorname{ri} C_1 \cap \operatorname{ri} C_2 = \phi$. Кроме того, C_1 и C_2 – выпуклые множества. По теореме 2.34 существует гиперплоскость H, собственно разделяющая C_1 и C_2. Гиперплоскость H невертикальна, поскольку вертикальная гиперплоскость разделяла бы множества $D(f_1)$ и $D(f_2)$, что противоречит условию теоремы. Таким образом, H имеет вид

$$H = H(a, \alpha) = \{(x, \xi)': \ (x - x^0)'a - \xi = \alpha\}.$$

Так как $(x^0; 0)' \in C_1 \cap C_2$, то $\alpha = 0$, и поэтому

$$(x - x^0)'a - \mu_1 f_1(x) + \mu_1 f_1(x^0) + (x - x^0)'y \leqslant 0 \quad \forall x \in D(f_1),$$

$$(x - x^0)'a - \mu_2 f_2(x^0) + \mu_2 f_2(x) \geqslant 0 \quad \forall x \in D(f_2),$$

и тем самым

$$a + y \in \mu_1 \partial_1(x^0), \quad -a \in \mu_2 \partial f_2(x^0).$$

Таким образом, $y \in \mu_1 \partial f_1(x^0) + \mu_2 \partial f_2(x^0)$. Отсюда следует

$$\partial f(x) \subseteq \mu_1 \partial f_1(x) + \mu_2 \partial f_2(x) \quad \forall x \in D(f).$$

2.2.6. Замкнутые функции. Рассматривая сопряженные функции, мы увидим (см. 3.2), что замкнутые надграфики выпуклых функций характеризуются некоторой операцией взятия оболочки. Предваряя этот материал, мы исследуем в этом разделе функции, надграфики которых являются замкнутыми множествами.

Назовем функцию f: $D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$, *замкнутой снизу*, если она полунепрерывна снизу на $D(f)$ и справедливо равенство $D(f) = \underline{D}$, где

$$\underline{D} = \{y: \ \varliminf_{\substack{x \to y \\ x \in D(f)}} f(x) < \infty\}.$$

Функция f: $D(f) \to \mathbf{R}$ называется *замкнутой сверху*, если $-f$ замкнута снизу.

Т е о р е м а 2.84. *Пусть* f: $D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$, – *произвольная функция. Множество* $\operatorname{epi} f$ *тогда и только тогда замкнуто, когда* f *замкнута снизу.*

Д о к а з а т е л ь с т в о. 1. Пусть множество $\operatorname{epi} f$ замкнуто. Тогда все предельные точки последовательностей из $\operatorname{epi} f$ являются элементами $\operatorname{epi} f$. Выберем последовательность $\{(x^i, f(x^i))'\} \subset \operatorname{epi} f$, для которой $\lim\limits_{i \to \infty} x^i = y$,

$$\lim_{i \to \infty} f(x^i) = \eta = \varliminf_{\substack{x \to y \\ x \in D(f)}} f(x) < \infty.$$

Имеем

$$-\infty < f(y) \leqslant \eta = \lim_{i \to \infty} f(x^i) = \underline{\lim_{\substack{x \to y \\ x \in D(f)}}} f(x) < \infty. \tag{2.100}$$

Отсюда следует, что функция $f = f(x)$ полунепрерывна снизу и $\underline{D} \subseteq D(f)$.

На основании определения нижнего предела

$$\underline{\lim_{\substack{x \to y \\ x \in D(f)}}} f(x) \leqslant f(y), \tag{2.101}$$

и поэтому $D(f) \subseteq \underline{D}$. Таким образом, $D(f) = \underline{D}$.

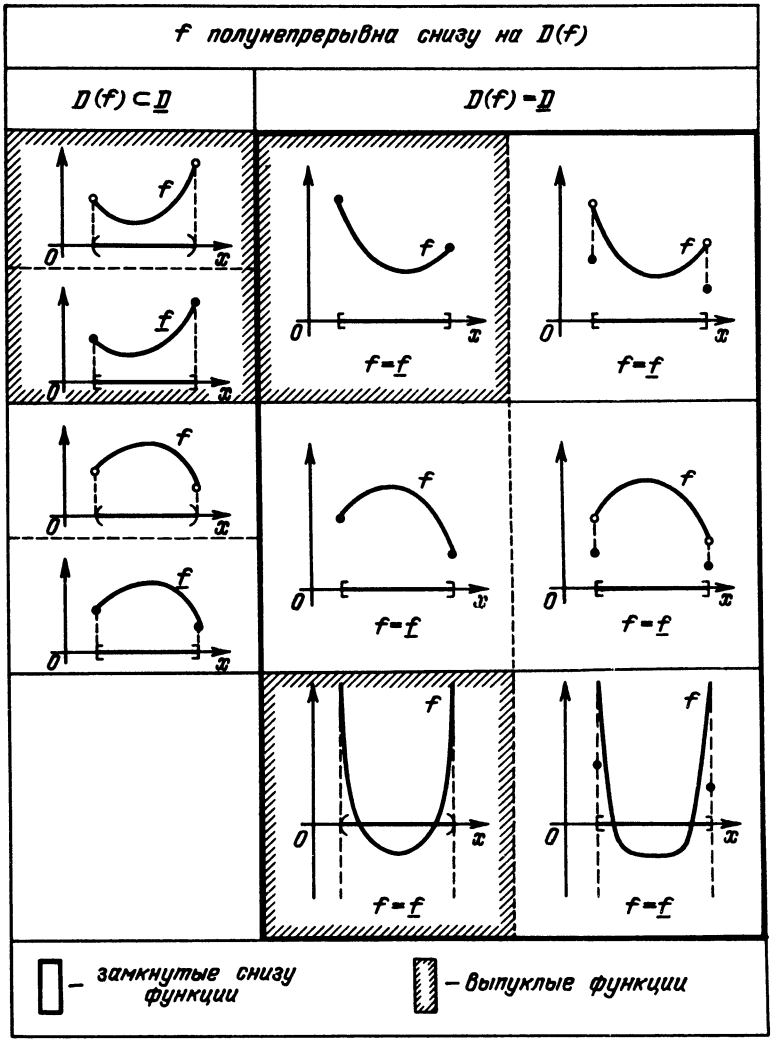


Рис. 2.19.

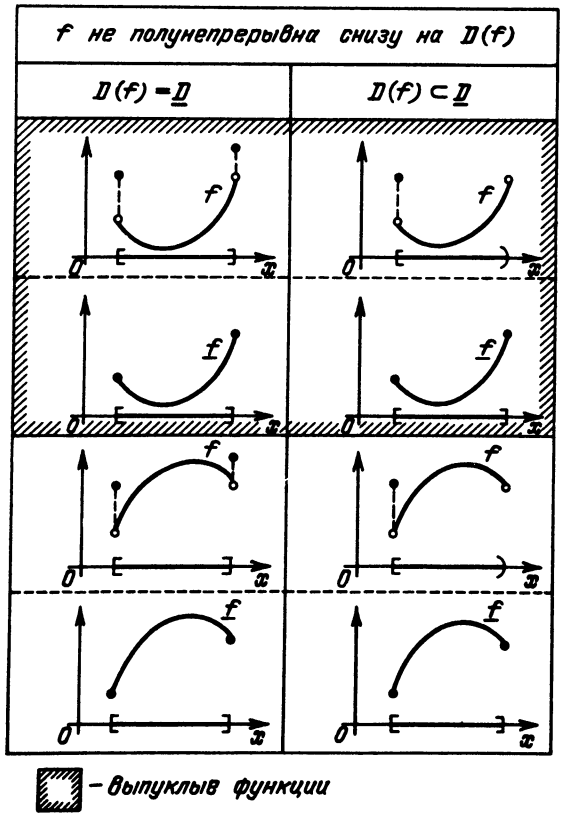


Рис. 2.20.

2. Пусть f – замкнутая снизу функция. Выберем произвольную последовательность

$$\{(x^i, \xi^i)'\} \subset \operatorname{epi} f, \quad \lim_{i \to \infty} x^i = y, \quad \lim_{i \to \infty} \xi^i = \eta. \tag{2.102}$$

Тогда $\xi^i \geqslant f(x^i)$ и

$$\eta = \lim_{i \to \infty} \xi^i = \varliminf_{x^i \to y} f(x^i) \geqslant \varliminf_{\substack{x \to y \\ x \in D(f)}} f(x) = f(y). \tag{2.103}$$

Итак, справедливо $(y, \eta)' \in \operatorname{epi} f$.

Как показывает следующий пример, замкнутая снизу выпуклая функция не обязана быть непрерывной.

Пример 2.23. Функция

$$f(x) = \begin{cases} x_1^2/x_2, & \text{если } x = (x_1, x_2)' \in \mathbf{R}^2, \quad x_2 \geqslant x_1^4, \quad x \neq 0, \\ 0, & \text{если } x = 0, \end{cases}$$

выпукла по теореме 2.74, поскольку для произвольного $x \in \mathbf{R}^2$, $x_2 > 0$, определитель матрицы Гессе этой функции равен нулю. Далее, для всех $x \in D(f)$ выполняется

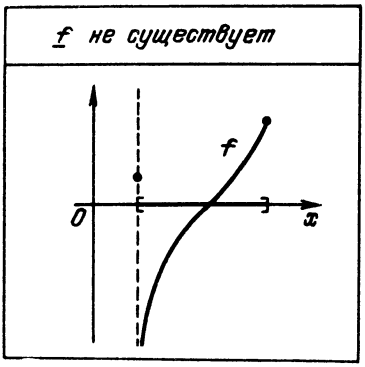


Рис. 2.21

$f(x) \geqslant 0$ и $\lim\limits_{(\xi_1, \xi_2)' \to (x_1, x_2)'} f(\xi_1, \xi_2) = f(x_1, x_2)$ $\forall (x_1, x_2)' \neq (0, 0)'$. Если двигаться в $D(f)$ вдоль прямой $x_2 = \alpha x_1$ ($\alpha > 0$) к началу координат, то получим $\lim\limits_{x_1 \to 0} f(x_1, \alpha x_1) = 0 = f(0)$. Так как кроме того, f непрерывна на $D(f) \setminus \{0\}$, то это означает, что функция f замкнута снизу. При приближении же к началу координат вдоль лежащего в $D(f)$ отрезка кубической параболы $x_2 = x_1^3$ функция f стремится к $+\infty$ в силу соотношения

$$\lim_{x_1 \to +0} f(x_1, x_1^3) = \lim_{x_1 \to +0} 1/x_1 = +\infty.$$

Если для функции f: $D(f) \to \mathbf{R}$ выполняется неравенство

$$\lim_{\substack{x \to y \\ x \in D(f)}} f(x) > -\infty \quad \forall y \in \overline{D(f)},$$

то можно определить $\underline{f}$ – *замыкание снизу* функции f (кратко: *замыкание* f) как

$$\underline{f}(y) = \lim_{\substack{x \to y \\ x \in D(f)}} f(x), \qquad y \in D(\underline{f}) = \{y: \lim_{\substack{x \to y \\ x \in D(f)}} f(x) < \infty\} = \overline{D(f)}.$$

Нетрудно увидеть, что $\underline{f}$ – всегда замкнутая снизу функция.

Если, в частности, f – замкнутая снизу функция, то $\underline{f} = f$.

Рис. 2.19 содержит примеры функций, которые полунепрерывны снизу и замкнуты снизу на множестве $D(f)$. На рис. 2.20 показаны функции, которые, равно как и их замыкания $\underline{f}$, не являются полунепрерывными снизу. При этом могут встретиться оба случая: и $D(f) = D(\underline{f})$, и $D(f) \subset D(\underline{f})$.

Для произвольной функции f: $D(f) \to \mathbf{R}$ не обязательно существует ее замыкание $\underline{f}$ (рис. 2.21).

Теорема 2.85. *Пусть* f: $D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$, – *функция, для которой*

$$\lim_{\substack{x \to y \\ x \in D(f)}} f(x) > -\infty \quad \forall y \in \overline{D(f)}.$$

Тогда

$$\overline{\operatorname{epi} f} = \operatorname{epi} \underline{f}.$$

Доказательство. 1. Покажем вначале справедливость включения $\overline{\operatorname{epi} f} \subset \operatorname{epi} \underline{f}$.

Пусть $(y, \eta)' \in \overline{\operatorname{epi} f}$. Тогда существует последовательность $(x^i, \eta^i)' \subset \operatorname{epi} f$ такая, что $(x^i, \eta^i)' \to (y, \eta)'$; при этом $\eta^i \geqslant f(x^i)$ $\forall i \in \mathbf{N}$. Тем самым

$$\eta \geqslant \lim_{\substack{x \to y \\ x \in D(f)}} f(x) = \underline{f}(y) \qquad \forall y \in \overline{D(f)},$$

и вследствие этого $(y, \eta)' \in \operatorname{epi} \underline{f}$.

2. Пусть $(y, \eta)' \in \operatorname{epi} \underline{f}$. Тогда с необходимостью $y \in \overline{D(f)}$ и

$$\eta \geqslant \underline{f}(y) = \varliminf_{\substack{x \to y \\ x \in D(f)}} f(x) > -\infty.$$

Поэтому существует последовательность $\{x^i\} \subset D(f)$, для которой $x^i \to y$ и $f(x^i) \to \underline{f}(y)$. Итак, $(y, \underline{f}(y))'$ – предельная точка множества $\operatorname{epi} f$, т.е. $(y, \underline{f}(y))' \in \overline{\operatorname{epi} f}$. Отсюда следует $\operatorname{epi} \underline{f} \subseteq \overline{\operatorname{epi} f}$, что вместе с доказанным ранее включением и дает утверждение теоремы.

Из теоремы 2.55 вытекает, что для выпуклой функции f: $D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$, множества уровня $N_\alpha = \{x \in D(f)$: $f(x) \leqslant \alpha\}$, $\alpha \in \mathbf{R}$, выпуклы.

Очевидно, для замкнутых снизу выпуклых функций утверждение теоремы 2.55 может быть усилено следующим образом.

Теорема 2.86. *Пусть $D(f) \subseteq \mathbf{R}^n$ – выпуклое множество и f: $D(f) \to \mathbf{R}$ – замкнутая снизу выпуклая функция. Тогда для произвольных $\alpha \in \mathbf{R}$ множества уровня N_α замкнуты и выпуклы.*

Исследуем теперь для замкнутой снизу выпуклой функции f: $D(f) \to \mathbf{R}$ вопрос об ограниченности множеств уровня N_α в зависимости от параметра α.

Теорема 2.87. *Пусть $D(f) \subseteq \mathbf{R}^n$ – выпуклое множество и f: $D(f) \to \mathbf{R}$ – замкнутая снизу выпуклая функция. Если для заданного $\alpha_0 \in \mathbf{R}$ множество уровня N_{α_0} не пусто и ограничено, то все множества N_α, $\alpha \in \mathbf{R}$, ограничены либо пусты.*

Доказательство. Очевидно, справедлива импликация $\alpha > \beta \Rightarrow N_\alpha \supseteq N_\beta$. Пусть множество N_β не ограничено. Тогда не ограничены при $\alpha' > \beta$ и множества $N_{\alpha'}$. Так как N_α замкнуты и выпуклы, то по теореме 2.13 существует элемент $z^0 \neq 0$ такой, что $x + S \subseteq N_\beta \quad \forall x \in N_\beta$, где $S = \{z$: $z = \lambda z^0$, $\lambda \in \mathbf{R}_+\}$.

Пусть теперь $\alpha'' < \beta$ и $N_{\alpha''} \neq \phi$. Выберем произвольный фиксированный вектор $x^0 \in N_{\alpha''} \subseteq N_\beta$. Для любых $(\mu_1, \mu_2)' \in \operatorname{ri} P_+^2$ и $\lambda \in \mathbf{R}_+$ имеем тогда

$$f(x^0 + \lambda z^0) = f\left(\mu_1 x^0 + \mu_2 \left(x^0 + \frac{\lambda}{\mu_2} z^0\right)\right) \leqslant$$

$$\leqslant \mu_1 f(x^0) + \mu_2 f\left(x^0 + \frac{\lambda}{\mu_2} z^0\right) \leqslant \mu_1 \alpha'' + (1 - \mu_1)\beta.$$

Отсюда при $\mu_1 \to 1 - 0$ следует

$$f(x^0 + \lambda z^0) \leqslant \alpha'' \quad \forall \alpha \in \mathbf{R}_+.$$

Итак, $x + S \subseteq N_{\alpha''}$, и поэтому множество $N_{\alpha''}$ тоже не ограничено.

2.2.7. Выпуклые функции f: $\mathbf{R}^n \to \overline{\mathbf{R}}$. Для изложения теории двойственности целесообразно рассмотреть также отображения f: $\mathbf{R}^n \to \overline{\mathbf{R}}$, где $\overline{\mathbf{R}} = \mathbf{R} \cup \{+\infty\} \cup \{-\infty\}$. Ниже мы кратко остановимся на выпуклых функциях f: $\mathbf{R}^n \to \overline{\mathbf{R}}$.

Функция $f\colon \mathbf{R}^n \to \bar{\mathbf{R}}$ называется *выпуклой*, если

$$\operatorname{epi} f = \{(x, \xi)' \in \mathbf{R}^{n+1}\colon \ \xi \geqslant f(x)\}$$

– выпуклое множество, т.е. если для любых $x^1, x^2 \in \mathbf{R}^n$, $\mu \geqslant f(x^1)$, $\nu \geqslant f(x^2)$ и $(\lambda_1, \lambda_2)' \in P_+^2$ выполняется

$$f(\lambda_1 x^1 + \lambda_2 x^2) \leqslant \lambda_1 \mu + \lambda_2 \nu.$$

Если не существует вектора $x \in \mathbf{R}^n$, для которого $f(x) = -\infty$, и если условиться, что

$$\alpha + \infty = \infty \quad \forall \alpha \in \mathbf{R}; \qquad \infty + \infty = \infty; \qquad \alpha \cdot \infty = -\infty \quad \forall \alpha < 0;$$

$$\alpha \cdot \infty = \infty \quad \forall \alpha > 0; \qquad \infty \cdot \infty = \infty; \qquad -\infty \cdot \infty = -\infty,$$

то, очевидно, функция $f\colon \mathbf{R}^n \to \bar{\mathbf{R}}$ тогда и только тогда выпукла, когда (по аналогии с п. 2.2.1) справедливо неравенство

$$f(\lambda_1 x^1 + \lambda_2 x^2) \leqslant \lambda_1 f(x^1) + \lambda_2 f(x^2)$$

для любых $x^1, x^2 \in \mathbf{R}^n$ и $(\lambda_1, \lambda_2)' \in \operatorname{ri} P_+^2$.

Функция $g\colon \mathbf{R}^n \to \bar{\mathbf{R}}$ называется *вогнутой*, если $f = -g$ выпукла. Для произвольной функции $f\colon \mathbf{R}^n \to \bar{\mathbf{R}}$ определим множество

$$\operatorname{dom} f = \{x\colon \ f(x) < \infty\},$$

которое назовем *эффективной областью*. Очевидно, что если f – выпуклая функция, то $\operatorname{dom} f$ является выпуклым множеством.

Далее, если для выпуклой функции $\operatorname{dom} f \neq \phi$ и $f(x) > -\infty \quad \forall x \in \mathbf{R}^n$, то f называется *собственно выпуклой*.

Конечным выпуклым функциям $\tilde{f}\colon D(\tilde{f}) \to \mathbf{R}$, $D(\tilde{f}) \subseteq \mathbf{R}^n$ можно поставить во взаимно однозначное соответствие собственно выпуклые функции по правилу

$$f(x) = \begin{cases} \tilde{f}(x), & \text{если} \quad x \in D(\tilde{f}), \\ +\infty, & \text{если} \quad x \in \mathbf{R}^n \setminus D(\tilde{f}). \end{cases} \tag{2.104}$$

Несобственно выпуклые функции обладают следующим свойством.

Т е о р е м а 2.88. *Пусть функция* $f\colon \mathbf{R}^n \to \bar{\mathbf{R}}$ *выпукла и* $f(x^0) = -\infty$. *Тогда*

$$f(x) = -\infty \quad \forall x \in \operatorname{ri}(\operatorname{dom} f).$$

Д о к а з а т е л ь с т в о. Пусть $f(x^0) = -\infty$ и $x \in \operatorname{ri}(\operatorname{dom} f)$. Тогда существует точка $x^1 \in \operatorname{dom} f$ такая, что $x = (1-\lambda)x^0 + \lambda x^1$, где $\lambda \in (0, 1)$. Таким образом, для $\lambda \in (0, 1)$ должно выполняться $f(x) \leqslant (1-\lambda)\mu + \lambda\nu$ при всех $\mu, \nu \in \mathbf{R}$ таких, что $\mu \geqslant f(x^0)$ и $\nu \geqslant f(x^1)$. Отсюда и из неравенства $f(x^1) < +\infty$ следует $f(x) = -\infty$.

Согласно теореме 2.88 несобственно выпуклая функция может быть конечной лишь в относительно граничных точках множества $\operatorname{dom} f$.

Назовем функцию $f\colon \mathbf{R}^n \to \bar{\mathbf{R}}$ *полунепрерывной снизу в точке* x^0, *если*

$$f(x^0) \leqslant \varliminf_{x \to x^0} f(x).$$

Функция $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ называется *полунепрерывной снизу* (на $\mathbf{R}^n$), если f полунепрерывна снизу в каждой точке $x^0 \in \mathbf{R}^n$.

С введенным в п. 2.2.3 понятием полунепрерывности функции в точке это понятие связано следующим образом. Пусть функция $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ образована из $\tilde{f}\colon D(\tilde{f}) \to \mathbf{R}$ согласно (2.104). Функция $\tilde{f}$ полунепрерывна снизу в точке $x^0 \in D(\tilde{f})$ тогда и только тогда, когда f полунепрерывна снизу в $x^0 \in \operatorname{dom} f = D(\tilde{f})$. Однако из полунепрерывности снизу функции $\tilde{f}$ на $D(\tilde{f})$ не обязательно следует полунепрерывность снизу f на $\mathbf{R}^n$.

Пример 2.24. Очевидно, что функция

$$f(x) = \begin{cases} x, & \text{если} \quad x > 0, \\ +\infty, & \text{если} \quad x \leqslant 0, \end{cases}$$

не является полунепрерывной снизу в точке $x^0 = 0$ (а потому и на R). Зато функция $\tilde{f}\colon D(\tilde{f}) \to \mathbf{R}$, $D(\tilde{f}) = \{x \in \mathbf{R}\colon x > 0\}$, $\tilde{f}(x) = x$, полунепрерывна снизу на $D(\tilde{f})$.

Теорема 2.89. *Собственно выпуклая функция* $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *полунепрерывна снизу на* $\mathbf{R}^n$ *тогда и только тогда, когда она замкнута снизу на множестве* $\operatorname{dom} f$.

Доказательство этой теоремы вытекает из теоремы 2.85 с учетом следующего предложения.

Теорема 2.90. *Следующие утверждения эквивалентны*:

1) *функция* $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *полунепрерывна снизу на* $\mathbf{R}^n$;

2) *множество уровня* $N_\alpha = \{x \in \mathbf{R}^n\colon f(x) \leqslant \alpha\}$ *замкнуто для любого* $\alpha \in \mathbf{R}$;

3) *множество* $\operatorname{epi} f$ *замкнуто.*

Доказательство. 1) $\Rightarrow$ 3). Пусть $\{(x^i, \mu^i)'\} \subset \operatorname{epi} f$ и $(x^i, \mu^i)' \to (x, \mu)'$. Требуется показать, что $(x, \mu)' \in \operatorname{epi} f$. В силу 1) $\mu^i \geqslant f(x^i)$ и $\varliminf\limits_{i \to \infty} f(x^i) \geqslant f(x)$, т.е. $\mu = \lim\limits_{i \to \infty} \mu^i \geqslant \varliminf\limits_{i \to \infty} f(x^i) \geqslant f(x)$.

1) $\Rightarrow$ 2). Пусть $\{x^i\} \in N_\alpha$ и $x^i \to x$. Справедливо неравенство $\alpha \geqslant f(x^i)$ $\forall i \in \mathbf{N}$, а потому и $\alpha \geqslant \varliminf\limits_{i \to \infty} f(x^i) \geqslant f(x)$, т.е. $x \in N_\alpha$.

2) $\Rightarrow$ 1). Пусть $\{x^i\} \subset N_\alpha$, $x^i \to x$, причем существует $\mu = \lim\limits_{i \to \infty} f(x^i)$. Для каждого $\alpha > \mu$ и достаточно большого i имеем $f(x^i) \leqslant \alpha$ и, следовательно,

$$x \in \overline{\{y\colon f(y) \leqslant \alpha\}} = \{y\colon f(y) \leqslant \alpha\}.$$

Поскольку число $\alpha > \mu$ выбрано произвольно, то $f(x) \leqslant \mu = \lim\limits_{i \to \infty} f(x^i)$, и потому выполняется 1).

3) $\Rightarrow$ 1). Доказательство вытекает из рассуждений в начале п. 2.2.6.

Будем говорить, что функция $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ имеет в точке x^0 *субградиент* y^0, если выполняется соотношение

$$f(x) \geqslant f(x^0) + y^{0\prime}(x - x^0) \qquad \forall x \in \mathbf{R}^n. \tag{2.105}$$

Множество всех субградиентов f в точке x^0 называется *субдифференциалом* функции f в точке x^0 и обозначается $\partial f(x^0)$. Если $f(x^0)$ – конечная

величина и $\partial f(x^0)$ не пусто, то f – собственная функция, поскольку она согласно (2.105) минорируется аффинной функцией.

Приведенные в п. 2.2.5 соотношения между субградиентом и производной по направлению выпуклой функции могут быть легко перенесены на функции вида f: $\mathbf{R}^n \to \bar{\mathbf{R}}$.

Если $\tilde{f}$: $D(\tilde{f}) \to \bar{\mathbf{R}}$ – конечная выпуклая функция, то производная по направлению $\tilde{f}^+(x^0; h)$ для некоторого $h \in K(D(\tilde{f}) - x^0)$ может принимать значение $-\infty$, т.е. (выпуклая) функция $f^+(x^0; h)$ переменной h отображает $K(D(\tilde{f}) - x^0)$ в множество $(\mathbf{R} \cup \{-\infty\}) \subseteq \bar{\mathbf{R}}$.

Другие утверждения о функциях f: $\mathbf{R}^n \to \bar{\mathbf{R}}$ можно найти, например, в [74, 35].

2.3. Обобщения выпуклых функций

Требование, чтобы функция была выпуклой или вогнутой, представляет собой довольно серьезное ограничение, которое в задачах, встречающихся на практике, не всегда выполняется. Поэтому понятие выпуклой функции обобщалось самыми различными способами. Мы рассмотрим некоторые такие обобщения.

В нашем изложении мы ограничимся обобщениями только выпуклых функций. Если f – определенная в дальнейшем обобщенная выпуклая функция, то $-f$ – соответствующая обобщенная вогнутая функция.

Если не оговаривается особо, то под f будем понимать функцию f: $\mathbf{R}^n \to \mathbf{R}$ или f: $D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$, $D(f) \neq \phi$.

Сначала мы обсудим обобщение выпуклых функций, суть которого состоит в следующем. Вначале определяется выпуклость функции в точке и лишь потом выпуклость на множестве и, кроме того, опускается требование выпуклости области определения. Мы будем говорить поэтому о выпуклости в общем смысле.

Функция f: $D(f) \to \mathbf{R}$ называется *выпуклой в общем смысле в точке* $\tilde{x} \in D(f)$ относительно множества $D(f)$, если выполняется соотношение

$$\left.\begin{array}{l} x \in D(f),\ (\lambda_1, \lambda_2)' \in P_+^2 \\ \lambda_1 \tilde{x} + \lambda_2 x \in D(f) \end{array}\right\} \Rightarrow f(\lambda_1 \tilde{x} + \lambda_2 x) \leqslant \lambda_1 f(\tilde{x}) + \lambda_2 f(x). \qquad (2.106)$$

Функция f называется *выпуклой в общем смысле на множестве* $D(f)$, если она в каждой точке $x \in D(f)$ выпукла в общем смысле относительно $D(f)$.

Из этого определения непосредственно получаем следующее утверждение.

Т е о р е м а 2.91. *Пусть* $D(f) \subseteq \mathbf{R}^n$ *– выпуклое множество. Функция* f: $D(f) \to \mathbf{R}$ *тогда и только тогда выпукла в общем смысле на множестве* $D(f)$, *когда*

$$\left.\begin{array}{l} x^1, x^2 \in D(f) \\ (\lambda_1, \lambda_2)' \in P_+^2 \end{array}\right\} \Rightarrow f(\lambda_1 x^1 + \lambda_2 x^2) \leqslant \lambda_1 f(x^1) + \lambda_2 f(x^2).$$

Из теоремы 2.91 легко усматривается связь с функциями, которые выпуклы на $D(f)$ в смысле данного в п. 2.2.1 определения. А именно, справедлива следующая теорема.

Т е о р е м а 2.92. *Если* $D(f) \subseteq \mathbf{R}^n$ *– выпуклое множество, то функция* f: $D(f) \to \mathbf{R}$ *тогда и только тогда выпукла на* $D(f)$, *когда она выпукла в общем смысле на множестве* $D(f)$.

Очевидно, что эквивалентность в теореме 2.92 не сохраняется, если множество $D(f)$ не является выпуклым. Так, хотя функция

$$f(x) = x^2, \qquad x \in D(f) = [-1, 1] \setminus \{0\},$$

и не выпукла (поскольку $D(f)$ – множество не выпуклое), она в то же время выпукла в общем смысле на $D(f)$. Вообще, если $D(f) \subseteq \mathbf{R}^n$ – выпуклое множество, f: $D(f) \to \mathbf{R}$ – выпуклая функция и X – произвольное подмножество $D(f)$, то f выпукла в общем смысле на множестве $D(f) \setminus X$. Таким образом, понятие выпуклой в общем смысле функции охватывает более широкий класс функций, нежели понятие выпуклой функции.

Функция f: $D(f) \to \mathbf{R}$ называется *строго выпуклой в общем смысле в точке* $\tilde{x} \in D(f)$ относительно множества $D(f)$, если выполняется соотношение

$$\left.\begin{array}{l} x \in D(f), \quad x \neq \tilde{x} \\ (\lambda_1, \lambda_2)' \in \operatorname{ri} P_+^2 \\ \lambda_1 \tilde{x} + \lambda_2 x \in D(f) \end{array}\right\} \Rightarrow f(\lambda_1 \tilde{x} + \lambda_2 x) < \lambda_1 f(\tilde{x}) + \lambda_2 f(x). \tag{2.107}$$

Для строго выпуклых в общем смысле функций имеют место утверждения, аналогичные теоремам 2.91 и 2.92.

Функция f: $D(f) \to \mathbf{R}$ называется *квазивыпуклой* на выпуклом множестве $D(f)$, если

$$\left.\begin{array}{l} x^1, x^2 \in D(f) \\ f(x^2) < f(x^1) \\ (\lambda_1, \lambda_2)' \in P_+^2 \end{array}\right\} \Rightarrow f(\lambda_1 x^1 + \lambda_2 x^2) \leqslant f(x^1). \tag{2.108}$$

Функция f: $D(f) \to \mathbf{R}$ называется *строго квазивыпуклой* на выпуклом множестве $D(f)$, если

$$\left.\begin{array}{l} x^1, x^2 \in D(f) \\ f(x^2) < f(x^1) \\ (\lambda_1, \lambda_2)' \in \operatorname{ri} P_+^2 \end{array}\right\} \Rightarrow f(\lambda_1 x^1 + \lambda_2 x^2) < f(x^1). \tag{2.109}$$

Пусть $D(f) \subseteq \mathbf{R}^n$ – открытое множество и f: $D(f) \to \mathbf{R}$ – дифференцируемая на $D(f)$ функция.

Функция f называется *псевдовыпуклой* на множестве $D(f)$, если

$$\left.\begin{array}{l} x^1, x^2 \in D(f) \\ f(x^2) < f(x^1) \end{array}\right\} \Rightarrow (x^2 - x^1)' \nabla f(x^1) < 0. \tag{2.110}$$

Функция f называется *строго псевдовыпуклой* на $D(f)$, если

$$\left.\begin{array}{l} x^1, x^2 \in D(f) \\ x^1 \neq x^2 \\ f(x^2) \leqslant f(x^1) \end{array}\right\} \Rightarrow (x^2 - x^1)' \nabla f(x^1) < 0. \tag{2.111}$$

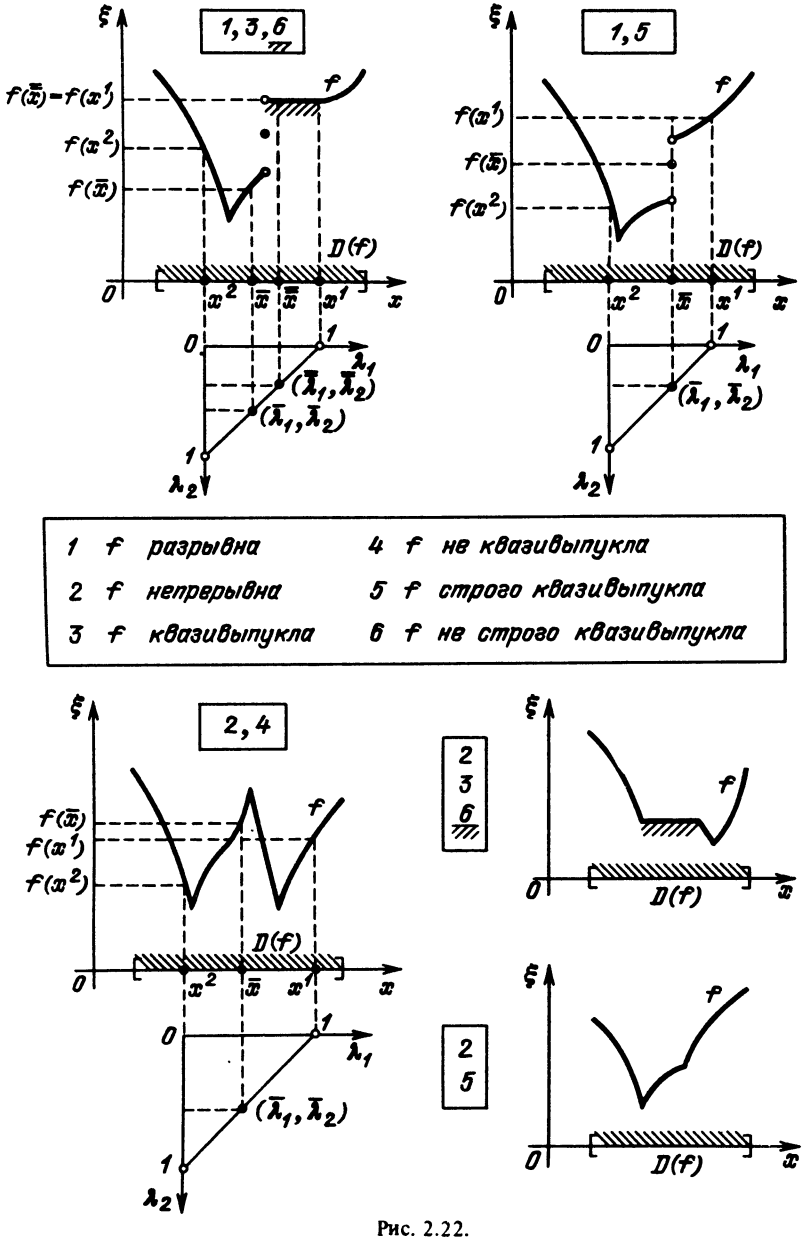
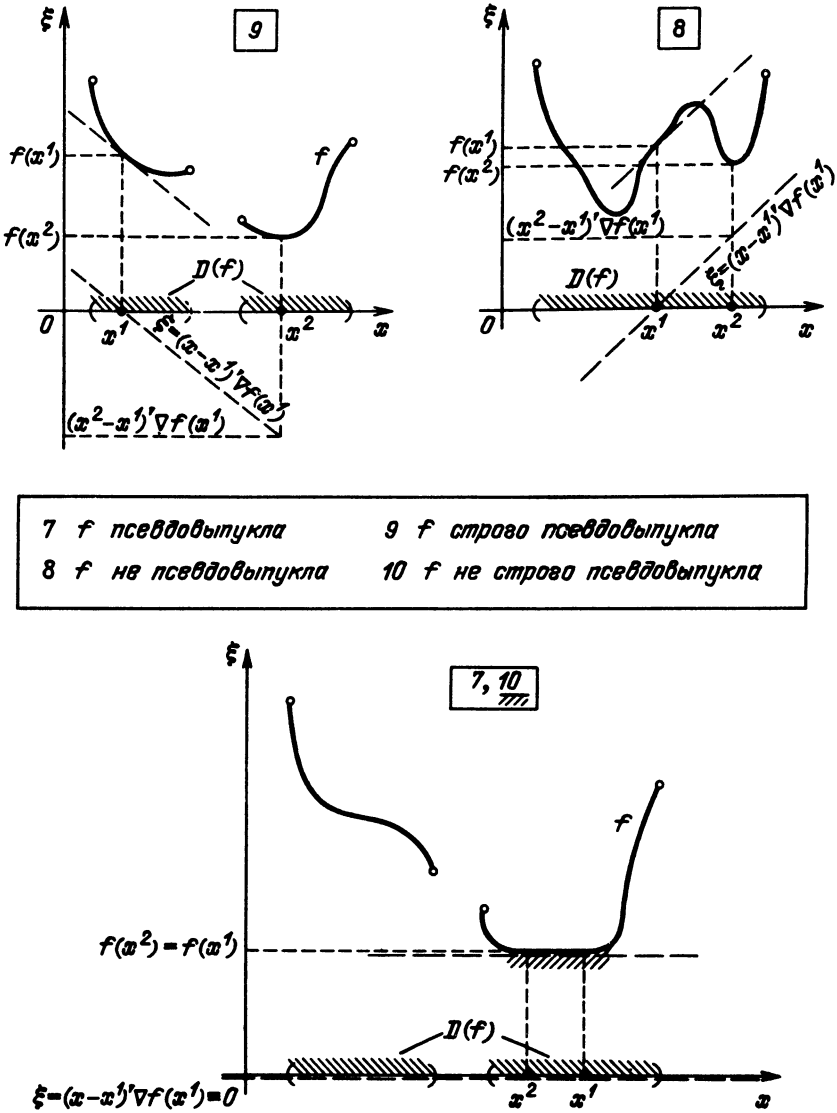


Рис. 2.22.

На рис. 2.22 и 2.23 представлены некоторые примеры только что введенных обобщений выпуклых функций. Приведем еще ряд примеров.

Пример 2.25. Функция

$$f_1(x) = \begin{cases} 0, & \text{если } x \in \mathbf{R} \setminus \mathbf{R}_+, \\ -x^n, & \text{если } x \in \mathbf{R}_+, \ n \in \mathbf{N}, \ n \geqslant 2, \end{cases}$$



7 f псевдовыпукла	9 f строго псевдовыпукла
8 f не псевдовыпукла	10 f не строго псевдовыпукла

Рис. 2.23.

квазивыпукла на множестве R, но не является ни строго квазивыпуклой, ни псевдовыпуклой, ни строго псевдовыпуклой, ни выпуклой.

Пример 2.26. Функция

$$f_2(x) = -x^n, \quad x \in \mathbf{R}_+, \quad n \in \mathbf{N}, \quad n \geqslant 2,$$

квазивыпукла и строго квазивыпукла на $\mathbf{R}_+$, но не является ни псевдовыпуклой, ни строго псевдовыпуклой, ни выпуклой.

Этот пример показывает также, что и нелинейная вогнутая функция может быть квазивыпуклой.

Пример 2.27. Функция

$$f_3(x) = -(x_1 + 1)^2 + 1, \quad x = (x_1, x_2)' \in P_+^2,$$

псевдовыпукла, строго квазивыпукла и квазивыпукла на множестве $\mathbf{R}_+^2$, но не является ни строго псевдовыпуклой, ни выпуклой.

Пример 2.28. Функция

$$f_4(x) = ax + b, \quad x \in \mathbf{R}, \quad a < 0,$$

выпукла, строго псевдовыпукла, псевдовыпукла, строго квазивыпукла и квазивыпукла на $\mathbf{R}$, но не строго выпукла.

Пример 2.29. Функция

$$f_5(x) = x^n, \quad x \in \mathbf{R}_+, \quad n \in \mathbf{N}, \quad n \geqslant 2,$$

строго выпукла, выпукла, строго псевдовыпукла, псевдовыпукла, строго квазивыпукла и квазивыпукла на множестве $\mathbf{R}_+$.

Непрерывные квазивыпуклые функции характеризуются следующей теоремой (доказательство см. в [16]).

Теорема 2.93. *Функция f: $D(f) \to \mathbf{R}$, непрерывная на выпуклом множестве $D(f) \subseteq \mathbf{R}^n$, тогда и только тогда квазивыпукла на $D(f)$, когда выполнено одно из следующих условий*:

1) $$\left.\begin{array}{l} x^1, x^2 \in D(f) \\ f(x^2) \leqslant f(x^1) \\ (\lambda_1, \lambda_2)' \in P_+^2 \end{array}\right\} \Rightarrow f(\lambda_1 x^1 + \lambda_2 x^2) \leqslant f(x^1); \qquad (2.112)$$

2) *множества уровня*

$$N_\alpha = \{x \in D(f): f(x) \leqslant \alpha\} \qquad (2.113)$$

выпуклы для любого $\alpha \in \mathbf{R}$;

3) *множества* $\operatorname{ri} N_\alpha$ *выпуклы для любого* $\alpha \in \mathbf{R}$. (2.114)

Для непрерывно дифференцируемых функций справедлив следующий факт (см. [31]).

Теорема 2.94. *Функция f: $D(f) \to \mathbf{R}$, непрерывно дифференцируемая на открытом выпуклом множестве $D(f) \subseteq \mathbf{R}^n$, тогда и только тогда квазивыпукла на $D(f)$, когда выполнено одно из следующих условий*:

1) $$\left.\begin{array}{l} x^1, x^2 \in D(f) \\ f(x^2) \leqslant f(x^1) \end{array}\right\} \Rightarrow (x^2 - x^1)' \nabla f(x^1) \leqslant 0; \qquad (2.115)$$

2) $$\left.\begin{array}{l} x^1, x^2 \in D(f) \\ f(x^2) < f(x^1) \end{array}\right\} \Rightarrow (x^2 - x^1)' \nabla f(x^1) \leqslant 0. \qquad (2.116)$$

Следующая теорема, которая вытекает непосредственно из (2.110), характеризует псевдовыпуклые функции.

Теорема 2.95. *Пусть $D(f) \subseteq \mathbf{R}^n$ – открытое множество и f: $D(f) \to \mathbf{R}$ дифференцируема на $D(f)$. Функция f тогда и только тогда псевдовыпукла на множестве $D(f)$, когда справедливо утверждение*

$$\left.\begin{array}{l} x^1, x^2 \in D(f) \\ (x^2 - x^1)' \nabla f(x^1) \geqslant 0 \end{array}\right\} \Rightarrow f(x^2) \geqslant f(x^1). \qquad (2.117)$$

Исследуем связь между приведенными выше обобщениями выпуклых функций и выпуклыми (а также строго выпуклыми) функциями.

Теорема 2.96. *Пусть* $D(f) \subseteq \mathbf{R}^n$ – *открытое выпуклое множество и функция* f: $D(f) \to \mathbf{R}$ *дифференцируема на* $D(f)$. *Если* f *выпукла на множестве* $D(f)$, *то она и псевдовыпукла на* $D(f)$.

Доказательство. Так как функция f выпукла, то по теореме 2.72 имеем

$$(x^2 - x^1)'\nabla f(x^1) \leqslant f(x^2) - f(x^1) \quad \forall x^1, x^2 \in D(f).$$

Поэтому соотношение $(x^2 - x^1)'\nabla f(x^1) \geqslant 0$ влечет неравенство $f(x^2) \geqslant f(x^1)$ и, следовательно, в силу теоремы 2.95 функция f псевдовыпукла на множестве $D(f)$.

Теорема 2.97. *Пусть* $D(f) \subseteq \mathbf{R}^n$ – *открытое выпуклое множество и функция* f: $D(f) \to \mathbf{R}$ *дифференцируема на* $D(f)$. *Если* f *строго выпукла на множестве* $D(f)$, *то она и строго псевдовыпукла на* $D(f)$.

Доказательство проводится аналогично доказательству теоремы 2.96 с учетом соответствующих строгих неравенств.

Следующая теорема не требует дифференцируемости функции.

Теорема 2.98. *Если функция* f: $D(f) \to \mathbf{R}$ *выпукла на выпуклом множестве* $D(f) \subseteq \mathbf{R}^n$, *то она и строго квазивыпукла на* $D(f)$.

Доказательство. Пусть $f(x^2) < f(x^1)$. В силу выпуклости f для $(\lambda_1, \lambda_2)' \in \operatorname{ri} P_+^2$ имеем

$$f(\lambda_1 x^1 + \lambda_2 x^2) \leqslant \lambda_1 f(x^1) + \lambda_2 f(x^2).$$

Поэтому неравенство $f(x^2) < f(x^1)$ влечет

$$f(\lambda_1 x^1 + \lambda_2 x^2) < f(x^1),$$

т.е. функция f строго квазивыпукла.

Обратные утверждения для теорем 2.96, 2.97 и 2.98 не имеют места, что показывают следующий пример и пример 2.26.

Пример 2.30. Функция $f(x) = x + x^3$, $x \in \mathbf{R}$, строго псевдовыпукла, а тем самым и псевдовыпукла, но не выпукла.

Между рассмотренными выше обобщениями выпуклых функций могут быть указаны следующие связи.

Теорема 2.99. *Пусть* $D(f) \subseteq \mathbf{R}^n$ – *открытое выпуклое множество и функция* f: $D(f) \to \mathbf{R}$ *дифференцируема на* $D(f)$.

Тогда:

1) *строгая псевдовыпуклость* f *на* $D(f)$ *влечет псевдовыпуклость* f *на* $D(f)$;

2) *псевдовыпуклость* f *на* $D(f)$ *влечет строгую квазивыпуклость* f *на* $D(f)$;

3) *строгая квазивыпуклость* f *на* $D(f)$ *влечет квазивыпуклость* f *на* $D(f)$.

Доказательство. 1) Указанная импликация следует непосредственно из определения.

2) Пусть $f(x^2) < f(x^1)$ и для $x^3 = \lambda_1 x^1 + \lambda_2 x^2$, где $(\lambda_1, \lambda_2)' \in \operatorname{ri} P_+^2$, выполняется $f(x^3) \geqslant f(x^1)$. Из этого предположения вытекает, что функция f достигает своего максимума на сегменте $[x^1, x^2]$ в точке $x^0 \in (x^1, x^2)$:

$$f(x^0) \geqslant f(x^3) \geqslant f(x^1) > f(x^2).$$

В силу псевдовыпуклости f отсюда следует

$$(x^2 - x^0)'\nabla f(x^0) < 0. \tag{2.118}$$

Положим $F(\lambda_1) = f(\lambda_1 x^1 + (1 - \lambda_1) x^2)$, $\lambda_1 \in (0, 1)$, тогда для некоторого $\lambda_1^0 \in (0, 1)$ имеем

$$f(x^0) = \max_{\lambda_1 \in (0,1)} F(\lambda_1) = F(\lambda_1^0).$$

Из необходимого условия оптимальности

$$\frac{dF(\lambda_1)}{d\lambda_1}\Bigg|_{\lambda_1 = \lambda_1^0} = (x^1 - x^2)'\nabla f(x^0) = 0$$

в силу $\lambda_1^0 \neq 0$ и $x^2 - x^0 = x^2 - \lambda_1^0 x^1 - (1 - \lambda_1^0) x^2 = \lambda_1^0 (x^2 - x_1)$ следует

$$(x^2 - x^0)'\nabla f(x^0) = 0, \tag{2.119}$$

что противоречит (2.118). Таким образом, $f(x^3) < f(x^1)$ для любого $x^3 \in (x^1, x^2)$.

3) Указанная импликация очевидна.

Нетрудно показать, что утверждение теоремы 2.99 нельзя обратить. Так, функция

$$f(x) = x^5, \quad x \in \mathbf{R},$$

строго квазивыпукла на $\mathbf{R}$, но не псевдовыпукла. Для $x_2 < 0$ из равенства $f'(0) = 0$, в противоречие с определением псевдовыпуклости, следует $f(x_2) < f(0)$ и $(x_2 - 0) f'(0) = 0$.

Как показывает следующая теорема, третья импликация в теореме 2.99 справедлива и без предположения о дифференцируемости функции f.

Т е о р е м а 2.100. *Если функция* f: $D(f) \to \mathbf{R}$ *строго квазивыпукла на выпуклом множестве* $D(f) \subseteq \mathbf{R}^n$, *то она и квазивыпукла на* $D(f)$.

Д о к а з а т е л ь с т в о. Для произвольного $(\lambda_1, \lambda_2)' \in$ ri P_+^2 тотчас же получаем, что соотношение (2.109) влечет за собой (2.108).

Следует указать, что в литературе нет единообразия в способах описания обобщений выпуклых функций. Так, можно найти определение квазивыпуклой функции с помощью соотношений (2.113) или (2.114), при этом не требуется непрерывность функции. В этом случае утверждение теоремы 2.100 не имеет места, что показывает следующий пример. Определенная на замкнутом интервале $[-1, 1] \subset \mathbf{R}$ функция

$$f(x) = \begin{cases} 1, & \text{если} \quad x = 0, \\ 0, & \text{если} \quad x \in [-1, 1], \ x \neq 0, \end{cases}$$

хотя и строго квазивыпукла в области определения, тем не менее, очевидно, не удовлетворяет ни (2.113), ни (2.114). Причиной этого является разрывность функции в точке $x = 0$.

Как известно, у выпуклой функции совпадают локальный и глобальный минимумы (см. теорему 2.62). Пример функции

$$f(x) = \begin{cases} 0, & \text{если} \quad x \in [-1, 0], \ x \in \mathbf{R}, \\ -x^2, & \text{если} \quad x \in (0, 1], \end{cases}$$

показывает, что это свойство не обязательно сохраняется у квазивыпуклой функции. Для строго квазивыпуклых функций (а в случае непрерывно дифференцируемых функций также для строго псевдовыпуклых и псевдовыпуклых функций) справедлива следующая теорема.

Теорема 2.101. *Пусть* $D(f) \subseteq \mathbf{R}^n$ *– выпуклое множество и* $f\colon D(f) \to \mathbf{R}$ *– строго квазивыпуклая на* $D(f)$ *функция. Любая точка локального минимума функции* f *на* $D(f)$ *является одновременно и точкой глобального минимума* f *на* $D(f)$.

Доказательство. Пусть функция f достигает в точке x^0 локального, но не глобального минимума на $D(f)$. Тогда найдется вектор $x^1 \in D(f)$ такой, что $f(x^1) < f(x^0)$. Отсюда в силу (2.109) следует

$$f(\lambda_1 x^0 + \lambda_2 x^1) < f(x^0) \qquad \forall (\lambda_1, \lambda_2)' \in \text{ri}\, P_+^2.$$

При $\lambda_2 \to +0$ получаем, что в любой окрестности x^0 существуют точки $x^2 = \lambda_1 x^0 + \lambda_2 x^1 \in D(f)$, для которых $f(x^2) < f(x^0)$. Это противоречит начальному предположению.

Теорема 2.75 показывает, что для выпуклой функции равенство градиента нулю является достаточным условием существования глобального минимума.

Для псевдовыпуклых, а тем самым и для строго псевдовыпуклых функций, имеет место соответствующая теорема.

Теорема 2.102. *Пусть* $D(f) \subseteq \mathbf{R}^n$ *– открытое множество и* $f\colon D(f) \to \mathbf{R}$ *– псевдовыпуклая на* $D(f)$ *функция. Если* $x^0 \in D(f)$ *и* $\nabla f(x^0) = 0$, *то* f *достигает в* x^0 *глобального минимума на* $D(f)$.

Доказательство. Для произвольного $x \in D(f)$ в силу $\nabla f(x^0) = 0$ имеем $(x - x^0)' \nabla f(x^0) = 0$, отсюда по теореме 2.95 вытекает $f(x) \geqslant f(x^0)$, т.е. f достигает в точке x^0 глобального минимума.

В заключение укажем, что имеется ряд других обобщений понятия выпуклой функции (см., например, [16, 17, 31]).

Для изложения методов нелинейного программирования (см. гл. 6) целесообразно остановиться вкратце еще на трех обобщениях выпуклых функций.

Пусть $D(f) \subseteq \mathbf{R}^n$ – непустое выпуклое множество. Функция $f\colon D(f) \to \mathbf{R}$ называется:

а) *явно квазивыпуклой* на $D(f)$, если

$$\left.\begin{array}{l} x^1, x^2 \in D(f) \\ f(x^2) \leqslant f(x^1) \\ (\lambda_1, \lambda_2)' \in P_+^2 \end{array}\right\} \Rightarrow f(\lambda_1 x^1 + \lambda_2 x^2) \leqslant f(x^1);$$

б) *строго явно квазивыпуклой* на $D(f)$, если

$$\left.\begin{array}{l} x^1, x^2 \in D(f), \quad x^1 \neq x^2 \\ f(x^2) \leqslant f(x^1) \\ (\lambda_1, \lambda_2)' \in \text{ri}\, P_+^2 \end{array}\right\} \Rightarrow f(\lambda_1 x^1 + \lambda_2 x^2) < f(x^1);$$

в) *почти выпуклой* на $D(f)$, если f явно квазивыпукла и строго квазивыпукла на $D(f)$.

Из этих определений непосредственно вытекают следующие утверждения о функции $f\colon D(f) \to \mathbf{R}$, где $D(f) \subseteq \mathbf{R}^n$ – выпуклое множество:

1) f строго выпукла на $D(f) \Rightarrow$
$\Rightarrow f$ строго явно квазивыпукла на $D(f)$,
$\Rightarrow f$ почти выпукла на $D(f)$,
$\Rightarrow f$ явно квазивыпукла на $D(f)$,
$\Rightarrow f$ квазивыпукла на $D(f)$;
2) f выпукла на $D(f) \Rightarrow f$ почти выпукла на $D(f)$;
3) f почти выпукла на $D(f) \Rightarrow \forall x^1, x^2 \in D(f)$, $\forall \lambda_1, \lambda_2 \in \operatorname{ri} P_+^2$:

$$f(x^2) \leqslant f(x^1) \Rightarrow f(\lambda_1 x^1 + \lambda_2 x^2) \leqslant f(x^1),$$

$$f(x^2) < f(x^1) \Rightarrow f(\lambda_1 x^1 + \lambda_2 x^2) < f(x^1).$$

2.4. Системы выпуклых неравенств

В этом параграфе мы рассмотрим системы выпуклых неравенств

$$g_i(x) \leqslant 0, \quad i \in I, \tag{2.120}$$

где I – конечное (соответственно произвольное) множество индексов и g_i: $C \to \mathbf{R}$, $C \subseteq \mathbf{R}^n$, – выпуклые функции.

Утверждения о разрешимости системы (2.120) могут быть получены при помощи теорем отделимости для выпуклых множеств. Их можно положить в основу теории глобальных критериев оптимальности (см. 4.3). Некоторые утверждения о разрешимости систем линейных неравенств получаются как частные случаи нижеследующих теорем.

Теорема 2.103. *Пусть $C \subseteq \mathbf{R}^n$ – выпуклое множество, f: $C \to \mathbf{R}^p$, g: $C \to \mathbf{R}^q$ – выпуклые вектор-функции и h: $C \to \mathbf{R}^r$ – аффинная вектор-функция. Справедливы утверждения*:

1) *если система*

$$f(x) < 0, \quad g(x) \leqslant 0, \quad h(x) = 0, \quad x \in C, \tag{2.121}$$

разрешима, то не существует вектора $(\tilde{u}, \tilde{v}, \tilde{w})' \in \mathbf{R}_+^p \times \mathbf{R}_+^q \times \mathbf{R}^r$ с $\tilde{u} \neq 0$, который бы был решением системы

$$u'f(x) + v'g(x) + w'h(x) \geqslant 0 \quad \forall x \in C; \tag{2.122}$$

2) *если система* (2.121) *неразрешима, то система* (2.122) *имеет нетривиальное решение $(\tilde{u}, \tilde{v}, \tilde{w})' \in \mathbf{R}_+^p \times \mathbf{R}_+^q \times \mathbf{R}^r$.*

Доказательство. 1. Если система (2.121) имеет решение x^0, то для $(\tilde{u}, \tilde{v}, \tilde{w})' \in \mathbf{R}_+^p \times \mathbf{R}_+^q \times \mathbf{R}^r$ с $\tilde{u} \neq 0$ должно выполняться неравенство

$$\tilde{u}'f(x^0) + \tilde{v}'g(x^0) + \tilde{w}'h(x^0) \leqslant \tilde{u}'f(x^0) < 0,$$

что противоречит (2.122).

2) Если система (2.121) не имеет решения, то выпуклые множества

$$C_1 = \{(y, z, t)' \in \mathbf{R}^p \times \mathbf{R}^q \times \mathbf{R}^r: \ \exists x \in C, \ f(x) < y, \ g(x) \leqslant z, \ h(x) = t\},$$

$$C_2 = \{(y, z, t)' \in \mathbf{R}^p \times \mathbf{R}^q \times \mathbf{R}^r: \ y \leqslant 0, \ z \leqslant 0, \ t = 0\}$$

не пересекаются. Тогда по теореме 2.34 существуют ненулевой вектор $(u, v, w)' \in \mathbf{R}^p \times \mathbf{R}^q \times \mathbf{R}^r$ и число $\alpha \in \mathbf{R}$ такие, что

$$y'u + z'v + t'w \geqslant \alpha \quad \forall (y, z, t)' \in C_1,$$

$$\alpha \geqslant y'u + z'v + t'w \quad \forall (y, z, t)' \in C_2.$$

Из второго неравенства следует, что

$$u \geqslant 0, \quad v \geqslant 0, \quad \alpha \geqslant 0.$$

Пусть $c \in \mathbf{R}^p$ и $c > 0$. Тогда из первого неравенства вытекает

$$u'f(x) + v'g(x) + w'h(x) \geqslant -u'c \qquad \forall x \in C.$$

Так как последнее неравенство выполняется для произвольного $c > 0$, то

$$u'f(x) + v'g(x) + w'h(x) \geqslant 0 \qquad \forall x \in C.$$

При доказательстве следующего утверждения нам понадобится теорема 2.103 для частного случая $h \equiv 0$.

Теорема 2.104. *Пусть $C \subseteq \mathbf{R}^n$ – выпуклое множество; $f\colon C \to \mathbf{R}^p$, $g\colon C \to \mathbf{R}^q$ – выпуклые вектор-функции и $h\colon \mathbf{R}^n \to \mathbf{R}^r$ – аффинная вектор-функция. Пусть существует вектор $\bar{x} \in \operatorname{ri} C$ такой, что $g(\bar{x}) < 0$, $h(\bar{x}) \leqslant 0$.*

Система неравенств

$$f(x) < 0, \quad g(x) \leqslant 0, \quad h(x) \leqslant 0, \quad x \in C, \tag{2.123}$$

неразрешима тогда и только тогда, когда существует вектор $(u, v, w)' \in \mathbf{R}^p_+ \times \mathbf{R}^q_+ \times \mathbf{R}^r$, $u \neq 0$, для которого

$$u'f(x) + v'g(x) + w'h(x) \geqslant 0 \qquad \forall x \in C. \tag{2.124}$$

Доказательство. Так же как при доказательстве теоремы 2.103, из разрешимости системы (2.123) заключаем, что (2.124) не имеет решений. Остается показать, что при предположениях теоремы 2.104 из неразрешимости (2.123) следует существование нетривиального решения системы (2.124) с $u \neq 0$.

Если система (2.123) не имеет решения, то выпуклые множества

$$C_1 = \{(y, z, t)' \in \mathbf{R}^p \times \mathbf{R}^q \times \mathbf{R}^r\colon \exists x \in C,\ f(x) < y,\ g(x) \leqslant z,\ h(x) = t\},$$

$$C_2 = \{(y, z, t)' \in \mathbf{R}^p \times \mathbf{R}^q \times \mathbf{R}^r\colon \ y \leqslant 0,\ z \leqslant 0,\ t \leqslant 0\}$$

не пересекаются. Применим следующую теорему отделимости (см. теорему 20.2 из [74]):

если C_1 – выпуклое множество, C_2 – многогранное выпуклое множество и $C_2 \cap \operatorname{ri} C_1 \neq \phi$, то существует гиперплоскость, которая собственно разделяет C_1 и C_2 и не содержит C_2.

Из этой теоремы для заданных множеств C_1 и C_2 следует существование вектора $(u, v, w)' \in \mathbf{R}^p \times \mathbf{R}^q \times \mathbf{R}^r$, $(u, v, w)' \neq 0$, и числа α, для которых

$$\begin{aligned} &u'y + v'z + t'w \geqslant \alpha \quad \forall (y, z, t)' \in C_1, \\ &\alpha \geqslant u'y + v'z + t'w \quad \forall (y, z, t)' \in C_2. \end{aligned} \tag{2.125}$$

Первое неравенство в (2.125) хотя бы для одного $(y, z, t)' \in C_1$ выполняется строго, т.е. найдется вектор $x^0 \in C$ такой, что

$$u'y + v'z + w'h(x^0) > \alpha, \quad f(x^0) < y, \quad g(x^0) \leqslant z. \tag{2.126}$$

Аналогично доказательству теоремы 2.103 заключаем, что $u \geqslant 0$, $v \geqslant 0$, $w \geqslant 0$, $\alpha \geqslant 0$ и

$$u'f(x) + v'g(x) + t'h(x) \geqslant 0 \quad \forall x \in C. \tag{2.127}$$

Допустим, $u = 0$. Тогда в силу существования вектора $\bar{x} \in \mathrm{ri}C$, удовлетворяющего неравенствам $g(\bar{x}) < 0$ и $h(\bar{x}) \leqslant 0$, из (2.127) немедленно следует $v = 0$. Поэтому, с одной стороны, $t' h(x) \geqslant 0$, а с другой, $h(\bar{x}) \leqslant 0$ для $\bar{x} \in \mathrm{ri}C$. Так как $t'h(x)$ – аффинная функция, то отсюда вытекает, что $t' h(x) = 0 \ \forall x \in C$. Но тогда из (2.126) при $u = 0$ и $v = 0$ получаем противоречие. Следовательно, $u \neq 0$.

Докажем теперь теорему о системе со сколь угодно большим числом выпуклых неравенств, из которой могут быть выведены важные следствия.

Пусть Z_1, Z_2 – множества индексов произвольной мощности.

Теорема 2.105. *Пусть $C \subseteq \mathbf{R}^n$ – непустое компактное выпуклое множество, $\{ g_i : C \to \mathbf{R} \}_{i \in Z_1}$ – семейство полунепрерывных снизу выпуклых функций и $\{ h_j : \mathbf{R}^n \to \mathbf{R} \}_{j \in Z_2}$ – семейство аффинных функций. Если система*

$$g_i(x) \leqslant 0, \quad i \in Z_1; \qquad h_j(x) = 0, \quad j \in Z_2; \qquad x \in C, \tag{2.128}$$

не имеет решения, то для некоторых конечных подсемейств

$$\{ g_{i_1}, \ldots, g_{i_m} \} \subset \{ g_i \}_{i \in Z_1}, \qquad \{ h_{j_1}, \ldots, h_{j_k} \} \subset \{ h_j \}_{j \in Z_2}$$

существует вектор $(\tilde{y}, \tilde{z})' \in \mathbf{R}_+^m \times \mathbf{R}^k$, $(\tilde{y}, \tilde{z})' \neq 0$, такой, что

$$\sum_{j=1}^{m} \tilde{y}_j g_{i_j}(x) + \sum_{i=1}^{k} \tilde{z}_i h_{j_i}(x) \geqslant 0 \qquad \forall x \in C. \tag{2.129}$$

Если $Z_2 = \phi$, то в (2.129) неравенство выполняется строго.

Доказательство. Если система (2.128) не разрешима, то не имеет решения и система

$$\begin{aligned} &g_i(x) \leqslant \epsilon, \quad i \in Z_1; \\ &h_j(x) = 0, \quad j \in Z_2; \qquad x \in C \quad \forall \epsilon > 0, \end{aligned} \tag{2.130}$$

ибо любое решение системы (2.130) является в то же время решением системы (2.128). Каждое из множеств

$$C_{ij\epsilon} = \{ x \in C: \ g_i(x) \leqslant \epsilon, \ h_j(x) = 0 \}$$

компактно как замкнутое подмножество компактного множества C, и, поскольку (2.130) не имеет решения, имеем

$$\bigcap_{\substack{(i,j) \in Z_1 \times Z_2 \\ \epsilon > 0}} C_{ij\epsilon} = \phi.$$

Применяя теорему Гейне – Бореля о покрытии к дополнениям множеств $C_{ij\epsilon}$, приходим к выводу, что можно выбрать конечное число $C_{ij\epsilon}$ таким образом, чтобы их пересечение было пусто. Это означает, что существуют индексы $i_1, \ldots, i_m \in Z_1$, $j_1, \ldots, j_k \in Z_2$ и действительные числа $\epsilon_1, \ldots, \epsilon_m > 0$ такие, что система

$$g_{i_j}(x) - \epsilon_j \leqslant 0, \quad j = 1, \ldots, m, \qquad h_{j_i}(x) = 0, \quad i = 1, \ldots, k, \quad x \in C,$$

не имеет решения.

Следовательно, по теореме 2.103 найдется вектор $(\tilde{y}, \tilde{z})' \in \mathbf{R}_+^m \times \mathbf{R}^k$, $(\tilde{y}, \tilde{z})' \neq 0$, для которого

$$\sum_{j=1}^{m} \tilde{y}_j g_{i_j}(x) + \sum_{i=1}^{k} \tilde{z}_i h_{j_i}(x) \geqslant \sum_{j=1}^{m} \tilde{y}_j \epsilon_j \qquad \forall x \in C.$$

Так как $\mu = \sum_{j=1}^{m} \tilde{y}_j \epsilon_i \geqslant 0$, то отсюда следует утверждение теоремы. Если $Z_2 = \phi$, то в силу $\tilde{y} \neq 0$ имеем даже $\mu > 0$.

Из теоремы 2.105 как частный случай вытекают утверждения о разрешимости систем линейных неравенств. В то время как теоремы о системах выпуклых неравенств мы будем использовать для вывода глобальных условий оптимальности, теоремы о системах линейных неравенств играют значительную роль в исследовании локальных условий оптимальности.

Т е о р е м а 2.106 (*теорема Моцкина об альтернативе*). *Пусть* A, B_1 *и* B_2 – *заданные матрицы порядка* $m_0 \times n$, $m_1 \times n$ *и* $m_2 \times n$ *соответственно. Из двух систем*

$$S_1: \quad Ax < 0, \quad B_1 x \leqslant 0, \quad B_2 x = 0, \tag{2.131}$$

$$S_2: \quad u'A + y'B_1 + z'B_2 = 0, \quad (u, y)' \geqslant 0, \quad u \neq 0, \tag{2.132}$$

разрешима ровно одна.

Д о к а з а т е л ь с т в о. Положим в теореме 2.104

$$C = \mathbf{R}^n, \quad f(x) = Ax, \quad g(x) = \epsilon, \quad \epsilon = \text{const},$$

$$\epsilon < 0; \quad h(x) = (B_1, B_2, -B_2)'x.$$

Все предположения теоремы 2.104 выполнены, и системе (2.124) соответствует система (2.131).

Легко видеть, что системы S_1 и S_2 не могут обе одновременно иметь решение. Предположим, что система S_1 не имеет решения. Тогда по теореме 2.104 найдется вектор $(\tilde{u}, \tilde{y}, \bar{w}, \bar{\bar{w}})' \in \mathbf{R}_+^{m_0} \times \mathbf{R}_+^{m_1} \times \mathbf{R}_+^{m_2} \times \mathbf{R}_+^{m_2}, \tilde{u} \neq 0$, для которого

$$\tilde{u}'Ax + \tilde{y}'B_1 x + (\bar{w} - \bar{\bar{w}})'B_2 x \geqslant 0 \quad \forall x \in \mathbf{R}^n. \tag{2.133}$$

Таким образом, линейная функция $l(x) = (\tilde{u}'A + \tilde{y}'B_1 + (\bar{w} - \bar{\bar{w}})'B_2)x$ неотрицательна для всех $x \in \mathbf{R}^n$. Но в этом случае

$$\tilde{u}'A + \tilde{y}'B_1 + (\bar{w} - \bar{\bar{w}})'B_2 = 0. \tag{2.134}$$

Пусть $\tilde{z} = \bar{w} - \bar{\bar{w}}$. Тогда $(\tilde{u}, \tilde{y}, \tilde{z})'$ – решение системы (2.132).

Положив в теореме 2.106 $B_1 = 0, B_2 = 0$, получим следующую теорему.

Т е о р е м а 2.107 (*теорема Жордана об альтернативе*). *Пусть* A – *матрица порядка* $m \times n$. *Из двух систем*

$$S_1: \quad Ax < 0,$$

$$S_2: \quad u'A = 0, \quad u \geqslant 0, \quad u \neq 0,$$

разрешима ровно одна.

Если в теореме 2.106 положить

$$A = -b', \quad b \in \mathbf{R}^n; \qquad B_1 = B, \quad B_2 = 0,$$

то после деления на $u > 0$ ($u \in \mathbf{R}$) получим теорему Фаркаша, которая была уже выведена другим путем в § 2.1.

Известная теорема Фаркаша для неоднородной системы линейных неравенств также может быть получена из теоремы 2.104.

Теорема 2.108 (*теорема Фаркаша для неоднородных систем*). *Пусть B – матрица порядка $m \times n$, $c \in \mathbf{R}^m$, $b \in \mathbf{R}^n$ и $\beta \in \mathbf{R}$. Из двух систем*

$$S_1 : Bx \leqslant c, \qquad b'x > \beta,$$

$$S_2 : B'y = b, \quad c'y \leqslant \beta, \quad y \geqslant 0$$

разрешима самое большее одна.

Доказательство. Положим в теореме 2.104 $f(x) = \beta - b'x$; $g(x) = \epsilon$, $\epsilon = \text{const}$, $\epsilon < 0$, $C = \mathbf{R}^n$, $h(x) = Bx - c$. Если существует $\bar{x} \in \mathbf{R}^n$ такой, что $B\bar{x} \leqslant c$, то по теореме 2.104 либо система S_1 имеет решение, либо существуют $\tilde{u} \in \mathbf{R}_+$, $\tilde{u} \neq 0$, $\tilde{v} \in \mathbf{R}_+$, $\tilde{w} \in \mathbf{R}_+^m$, для которых

$$\tilde{u}(\beta - b'x) + \tilde{w}'(Bx - c) = -\tilde{v}\epsilon \geqslant 0 \qquad \forall\, x \in \mathbf{R}^n. \tag{2.135}$$

Отсюда, положив $\tilde{y} = \tilde{w}/\tilde{u}$, получим

$$(\beta - \tilde{y}'c) + (\tilde{y}'B - b')x \geqslant 0 \qquad \forall\, x \in \mathbf{R}^n. \tag{2.136}$$

Но это возможно, лишь если одновременно $B'\tilde{y} = b$ и $c'\tilde{y} \leqslant \beta$. Следовательно, система S_2 разрешима.

Наконец, с помощью простых примеров можно показать, что как система $Bx \leqslant c$, так и $B'y = b$, $y \geqslant 0$, не обязательно имеют решения.

Перечень теорем об альтернативе для систем линейных неравенств, которые могут быть получены описанным выше или подобным способом, можно найти, например, в [59].

3.1. Сопряженные множества

Теория сопряженных функций, основы которой были заложены Фенхелем [26], образует важный подход к теории двойственности в нелинейном программировании. Исходным пунктом для получения теорем двойственности являются при этом двойственные друг другу задачи описания выпуклых замкнутых множеств с помощью точечного определения и через совокупность своих опорных гиперплоскостей.

Очевидно, что задача определения функции f множеством опорных к $\operatorname{epi} f$ ($\operatorname{hypo} f$) гиперплоскостей имеет смысл для выпуклых (вогнутых) функций. Более того, при ближайшем рассмотрении оказывается, что, с одной стороны, это "двойственное" описание f может привести к интересным утверждениям и для невыпуклых (невогнутых) функций, а с другой, – не каждая выпуклая (вогнутая) функция может быть описана подобным образом.

Вначале мы введем для произвольных множеств $M \subseteq \mathbf{R}^n \times \mathbf{R} = \mathbf{R}^{n+1}$ понятие сопряженного множества. Множества

$$M^c = \{ (y, w)' \in \mathbf{R}^{n+1} : y'x - z \leqslant w \quad \forall (x, z)' \in M \}, \tag{3.1}$$

$$M_c = \{ (y, w)' \in \mathbf{R}^{n+1} : y'x - z \geqslant w \quad \forall (x, z)' \in M \} \tag{3.2}$$

называются *множествами, сопряженными к M сверху и снизу* соответственно.

Для $M = \mathbf{R}^{n+1}$ имеем $M^c = M_c = \phi$. Если $M = \phi$, то условимся, что $M^c = M_c = \mathbf{R}^{n+1}$.

Так как в последующем рассматриваются преимущественно сопряженные сверху множества, то там, где это не приводит к недоразумениям, мы будем их кратко называть *сопряженными множествами.*

Приведем геометрическую интерпретацию множества M^c. Назовем ось z пространства точек $(x, z)'$ вертикальной осью. Любая точка $(y, w)' \in M^c$ обладает следующим свойством: порожденное невертикальной гиперплоскостью

$$H(y, w) = \{ (x, z)' \in \mathbf{R}^{n+1} : y'x - z = w \}$$

полупространство

$$H_-(y, w) = \{ (x, z)' \in \mathbf{R}^{n+1} : y'x - z \leqslant w \}$$

содержит множество M. По определению множества M^c

$$M \subseteq H_-(y, w) \iff (y, w)' \in M^c. \tag{3.3}$$

Отсюда следует

$$M \subseteq \bigcap_{(y, w)' \in M^c} H_-(y, w).$$

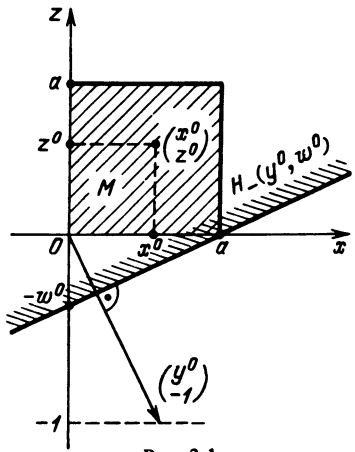


Рис. 3.1.

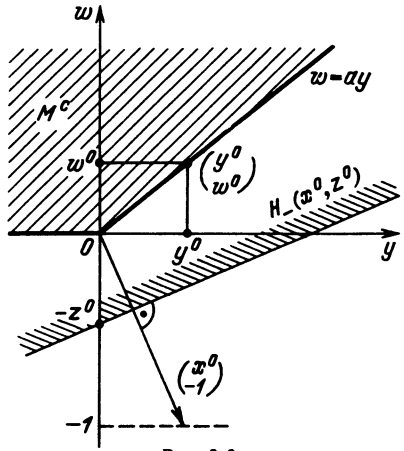


Рис. 3.2.

Поскольку

$$M^c = \{ (y, w)' \in \mathbf{R}^{n+1} : \sup_{(x, z)' \in M} (y'x - z) \leqslant w \},$$

то для описания множества M^c достаточно уже знания совокупности всех невертикальных опорных к множеству M гиперплоскостей

$$H(y, b) = \{ (x, z)' \in \mathbf{R}^{n+1} : y'x - z = b \},$$

где $b = \sup_{(x, z)' \in M} (y'x - z) < \infty$.

Пример 3.1. Для множества $M = \{ (x, z)' : x, z \in [0, a]; a \geqslant 0 \} \subseteq \mathbf{R}^2$ (рис. 3.1) имеем (рис. 3.2)

$$M^c = \{ (y, w)' : yx - z \leqslant w \quad \forall (x, z)' \in M \} =$$

$$= \{ (y, w)' : y < 0, w \geqslant 0 \} \cup \{ (y, w)' : y \geqslant 0, w \geqslant ay \},$$

$$M_c = \{ (y, w)' : yx - z \geqslant w \quad \forall (x, z)' \in M \} =$$

$$= \{ (y, w)' : ya - a \geqslant w, y < 0 \} \cup \{ (y, w)' : w = -a, y \geqslant 0 \}.$$

Вообще говоря, совершенно различные по своей природе множества могут обладать одними и теми же сопряженными множествами. Так, легко проверить, что для множеств $\widetilde{M}_1 = \operatorname{int} M$ и $\widetilde{M}_2 = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ a \end{pmatrix}, \begin{pmatrix} a \\ 0 \end{pmatrix}, \begin{pmatrix} a \\ a \end{pmatrix} \right\}$ выполняется равенство $\widetilde{M}_1^c = \widetilde{M}_2^c = M^c$. Непосредственно из определения M^c и M_c следует, что

$$M_1 \subseteq M_2 \Rightarrow M_2^c \subseteq M_1^c \tag{3.4}$$

и соответственно

$$M_1 \subseteq M_2 \Rightarrow (M_2)_c \subseteq (M_1)_c \tag{3.5}$$

Пример 3.2. Для изображенных на рис. 3.3 множеств M_1, M_2 построены множества, сопряженные сверху (рис. 3.4) и снизу (рис. 3.5). Легко видеть справедливость соотношений (3.4) и (3.5).

Пусть $(x, z)' \in \mathbf{R}^{n+1}$ и

$$H_-(x, z) = \{ (y, w)' \in \mathbf{R}^{n+1} : y'x - w \leqslant z \}.$$

Из определения множества M^c следует, что

$$(y, w)' \in M^c \iff y'x - w \leqslant z \quad \forall (x, z)' \in M,$$

$$\iff (y, w)' \in \bigcap_{(x,z)' \in M} H_-(x, z),$$

и тем самым

$$M^c = \bigcap_{(x,z)' \in M} H_-(x, z). \tag{3.6}$$

Как пересечение замкнутых полупространств M^c – замкнутое выпуклое множество.

Операцию образования сопряженного (сверху или снизу) множества можно, в частности, применить к множеству, которое само является сопряженным к заданному множеству M. Определим

$$M^{cc} = (M^c)^c = \bigcap_{(y,w)' \in M^c} H_-(y, w) \tag{3.7}$$

и назовем M^{cc} *множеством, дважды сопряженным* (сверху) к M. Очевидно M^{cc} – замкнутое множество. Аналогично назовем

$$M^{ccc} = (M^{cc})^c = ((M^c)^c)^c = (M^c)^{cc}$$

множеством, трижды сопряженным (сверху) к M, и т.д. Из (3.7) следует, что

$$(x, z)' \in M^{cc} \iff y'x - z \leqslant w \quad \forall (y, w)' \in M^c.$$

В силу (3.1) соотношение $y'x - z \leqslant w \quad \forall (y, w)' \in M^c$ выполняется, в частности, для всех $(x, z)' \in M$, т.е.

$$M \subseteq M^{cc}. \tag{3.8}$$

Замечание после примера 3.1 показывает, что равенство $M = M^{cc}$, вообще говоря, не имеет места.

Однако, применяя (3.7) и (3.8), из (3.4) получим

$$M^{ccc} = (M^{cc})^c = M^c, \tag{3.9}$$

потому что, с одной стороны, выполняется соотношение $M \subseteq M^{cc} \Rightarrow M^c \supseteq \supseteq (M^{cc})^c = (M^c)^{cc} = M^{ccc}$, а с другой, – $M^c \subseteq (M^c)^{cc}$.

Из свойств (3.4), (3.8) и (3.9) следует, что соответствие между M^{cc} и M является некоторой операцией взятия оболочки M, т.е. справедливы

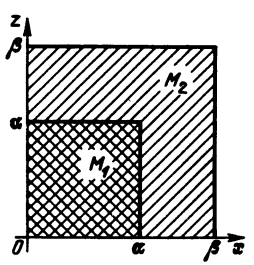


Рис. 3.3.

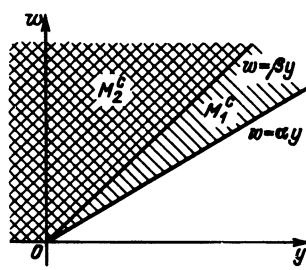


Рис. 3.4.

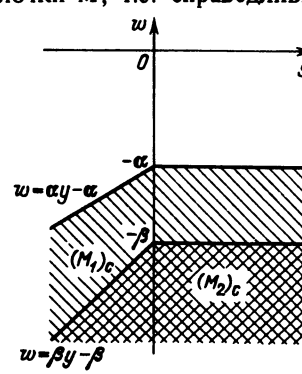


Рис. 3.5.

соотношения

1) $M \subseteq M^{cc}$,

2) $M_1 \subseteq M_2 \Rightarrow M_1^{cc} \subseteq M_2^{cc}$,

3) $M^{cc} = (M^{cc})^{cc}$.

Ниже будет показано, какие множества характеризуются операцией взятия оболочки M в том смысле, что выполняется равенство $M^{cc} = M$. Это те надграфики $M \subseteq \mathbf{R}^{n+1}$, для которых наряду с точечным определением возможно и двойственное описание как пересечения (замкнутых) полупространств с невертикальными образующими гиперплоскостями.

Теорема 3.1. *Пусть $M \subseteq \mathbf{R}^{n+1}$ – произвольное множество такое, что $M^c \neq \phi$. Тогда M^c является замкнутым выпуклым надграфиком некоторой функции g: $D(g) \to \mathbf{R}$, $D(g) \subseteq \mathbf{R}^n$.*

Доказательство. Прежде всего заметим, что имеет место соотношение

$$(y^0, w^0)' \in M^c \Rightarrow (y^0, w)' \in M^c \quad \forall\, w \geqslant w^0. \tag{3.10}$$

Множество M^c не содержит вертикальных прямых, так как иначе существовал бы вектор $\bar{y}^0 \in \mathbf{R}^n$ такой, что $(\bar{y}^0, w)' \in M^c$ для любого $w < 0$, а это противоречит (3.3). Поэтому для любого $(y^0, w^0)' \in M^c$ имеем $\tilde{w}^0 = \inf_{(y^0, w)' \in M^c} w > -\infty$.

Поскольку M^c – замкнутое выпуклое множество, то $(y^0, \tilde{w}^0)' \in M^c$. Положив $g(y^0) = \tilde{w}^0$ для $y^0 \in \Pi_{\mathbf{R}^n}(M^c)$, получим функцию g: $D(g) = \Pi_{\mathbf{R}^n}(M^c) \to \mathbf{R}$, у которой $\operatorname{epi} g = M^c$. Тем самым теорема доказана.

Согласно теоремам 2.56 и 2.84 надграфик $\operatorname{epi} g$ функции g: $D(g) \to \mathbf{R}$, $D(g) \subseteq \mathbf{R}^n$, тогда и только тогда является замкнутым выпуклым множеством, когда g выпукла и замкнута снизу. Если M – непустое множество, $M^c \neq \phi$ и $M^{cc} = M$, то по теореме 3.1 множество $M = M^{cc} = (M^c)^c$ является надграфиком некоторой замкнутой снизу выпуклой функции.

Как мы покажем ниже, справедливо и обратное утверждение, а именно, имеет место следующая теорема.

Теорема 3.2. *Непустое множество $M \subset \mathbf{R}^{n+1}$ тогда и только тогда является надграфиком некоторой замкнутой снизу выпуклой функции, когда $M^{cc} = M$ и $M^c \neq \phi$.*

Доказательство. В силу сказанного выше достаточно показать, что для любого множества $M \subseteq \mathbf{R}^{n+1}$, которое является награфиком некоторой замкнутой снизу выпуклой функции, выполняется соотношение $M^{cc} = M$.

Как мы знаем, справедливо соотношение $M \subseteq M^{cc}$. Согласно теоремам 2.56 и 2.84 M как надграфик замкнутой снизу выпуклой функции является замкнутым выпуклым множеством, не содержащим вертикальных прямых. В силу теоремы 2.77 существует невертикальная опорная к M гиперплоскость $H(a, \alpha)$ такая, что $M \subseteq H_-(a, \alpha)$.

Пусть теперь $(x^0, \xi^0)' \notin M$. Если $a'x^0 - \xi^0 > \alpha$, то $(x^0, \xi^0)' \notin H_-(a, \alpha)$ и тем самым, в силу (3.3) и (3.7), $(x^0, \xi^0)' \notin M^{cc}$.

Мы завершим доказательство, если покажем, что ни одна точка $(x^0, \xi^0)' \notin M$, для которой $a'x^0 - \xi^0 \leqslant \alpha$, не принадлежит M^{cc}. Пусть $(x^0, \xi^0)' \notin M$ и $a'x^0 - \xi^0 \leqslant \alpha$. Для множеств $C_1 = M$ и $C_2 = \{(x^0, \xi^0)'\}$ по теореме 2.38

существует гиперплоскость $\widetilde{H} = \{(x, \xi)' \in \mathbf{R}^{n+1}: b'x - \gamma\xi = \beta\}$ такая, что

$$b'x - \gamma\xi \leqslant \beta < b'x^0 - \gamma\xi^0 \qquad \forall (x, \xi)' \in M. \tag{3.11}$$

Если мы выберем $x \in D(f)$, то $(x, \xi)' \in M$ для всех ξ таких, что $\xi \geqslant f(x)$. Отсюда и из левого неравенства в (3.11) получаем $\gamma \geqslant 0$. Для $\gamma > 0$ имеем

$$\begin{pmatrix} x^0 \\ \xi^0 \end{pmatrix} \notin H_-\left(\frac{b}{\gamma}, \frac{\beta}{\gamma}\right), \qquad M \subseteq H_-\left(\frac{b}{\gamma}, \frac{\beta}{\gamma}\right),$$

и, следовательно, $(x^0, \xi^0)' \notin M^{cc}$.

Если же $\gamma = 0$, то из $H(a, \alpha)$ и $\widetilde{H}$ можно построить невертикальную гиперплоскость, строго разделяющую C_1 и C_2. Рассмотрим два случая.

Случай 1. $a'x^0 - \xi^0 = \alpha$.

Тогда в силу (3.11), (3.12) и включения $M \subseteq H_-(a, \alpha)$ имеем

$$(a + b)'x - \xi \leqslant \alpha + \beta < (a + b)'x^0 - \xi^0 \qquad \forall (x, \xi)' \in M,$$

т.е.

$$M \subseteq H_-(a + b, \alpha + \beta), \qquad (x^0, \xi^0)' \notin H_-(a + b, \alpha + \beta).$$

Случай 2. $a'x^0 - \xi^0 < \alpha$. Тогда для любого ρ такого, что

$$0 < \frac{\alpha + \xi^0 - a'x}{b'x^0 - \beta} < \rho, \tag{3.13}$$

согласно (3.11) и включению $M \subseteq H_-(a, \alpha)$ имеет место неравенство

$$(a + \rho b)'x - \xi \leqslant \alpha + \rho\beta < (a + \rho b)'x^0 - \xi^0 \qquad \forall (x, \xi)' \in M,$$

т.е. $M \subseteq H_-(a + \rho b, \alpha + \rho\beta)$ и $(x^0, \xi^0)' \notin H_-(a + \rho b, \alpha + \rho\beta)$. Тем самым показано, что ни одна точка $(x^0, \xi^0)' \notin M$ не принадлежит M^{cc}. Таким образом, $M = M^{cc}$.

Из предыдущих рассуждений вытекает следующее утверждение об операции взятия оболочки M.

Теорема 3.3. *Пусть множества $M \subseteq \mathbf{R}^{n+1}$ и M^c непусты. Тогда M^{cc} – "наименьший" надграфик замкнутой снизу выпуклой функции g: $D(g) \to \mathbf{R}$, $D(g) \subseteq \mathbf{R}^n$, со свойством $M^{cc} = \operatorname{epi} g \supseteq M$.*

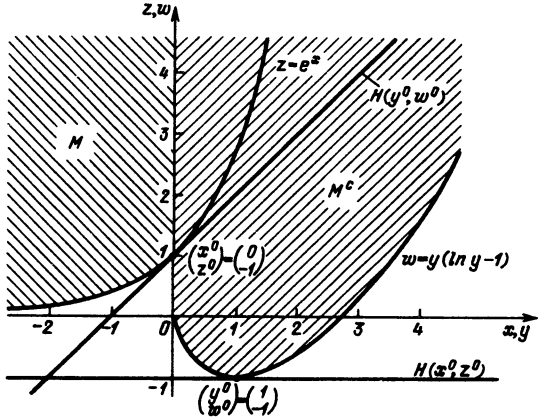


Рис. 3.6.

Доказательство. Пусть C – пересечение всех надграфиков замкнутых снизу выпуклых функций g: $D(g) \to \mathbf{R}, D(g) \subseteq \mathbf{R}^n$, со свойством $\operatorname{epi} g \supseteq M$. Так как M^{cc} – замкнутый выпуклый надграфик некоторой функции и $M \subseteq M^{cc}$, то, с одной стороны, имеем $C \subseteq M^{cc}$ а с другой, – в силу $C \supseteq M$, выполняется $C = (C)^{cc} \supseteq M^{cc}$.

Следующая теорема содержит утверждение об опорных точках гиперплоскостей, опорных к M и M^c (рис. 3.6).

Теорема 3.4. (*теорема о симметрии*). *Пусть $M \subseteq \mathbf{R}^{n+1}$ – произвольное непустое множество. Если $(x^0, z^0)'$ – опорная точка опорной к M гиперплоскости $H(y^0, w^0)$ такая, что $M \subseteq H_-(y^0, w^0)$, то $H(x^0, z^0)$ – опорная к M^c гиперплоскость, а $(y^0, w^0)'$ – опорная точка этой гиперплоскости.*

Доказательство. Учитывая условия теоремы, имеем

$$(x^0, z^0)' \in M \cap H(y^0, w^0) \Leftrightarrow y^{0\prime} x^0 - z^0 = w^0, \tag{3.14}$$

$$M \subseteq H_-(y^0, w^0). \tag{3.15}$$

Из (3.3) и (3.15) прежде всего следует $(y^0, w^0)' \in M^c$. Из (3.6) и (3.14) вытекает $M^c \subseteq H_-(x^0, z^0)$. В силу того, что

$$y^{0\prime} x^0 - z^0 = w^0 \Leftrightarrow y^{0\prime} x^0 - w^0 = z^0,$$

очевидно, выполняется соотношение

$$(x^0, z^0)' \in H(y^0, w^0) \Leftrightarrow (y^0, w^0)' \in H(x^0, z^0). \tag{3.16}$$

Поэтому в итоге имеем

$$(y^0, w^0)' \in M^c \cap H(x^0, z^0), \qquad M^c \subseteq H_-(x^0, z^0),$$

т.е. $(y^0, w^0)'$ – опорная точка гиперплоскости $H(x^0, z^0)$, опорной к множеству M^c.

В дальнейшем рассматриваются лишь такие множества $M \subseteq \mathbf{R}^{n+1}$, которые являются надграфиком или подграфиком некоторой функции f: $D(f) \to \mathbf{R}, D(f) \subseteq \mathbf{R}^n$.

3.2. Сопряженные функции

Для функции f: $D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$, имеем

$$\operatorname{epi} f = \{(x, z)' : \ x \in D(f), \ z \geqslant f(x)\},$$

$$(\operatorname{epi} f)^c = \{(y, w)' : \ y'x - z \leqslant w \quad \forall (x, z)' \in \operatorname{epi} f\} =$$

$$= \{(y, w)' : \ y'x - f(x) \leqslant w \quad \forall x \in D(f)\} =$$

$$= \{(y, w)' : \ \sup_{x \in D(f)} \{y'x - f(x)\} < \infty, \quad w \geqslant \sup_{x \in D(f)} \{y'x - f(x)\}\}.$$

Такое представление множества $(\operatorname{epi} f)^c$ позволяет ввести определение сопряженных функций. С этой целью для произвольной функции f: $D(f) \to \mathbf{R}, D(f) \subseteq \mathbf{R}^n$, введем множества

$$D(f^c) = \{y \in \mathbf{R}^n : \ \sup_{x \in D(f)} \{y'x - f(x)\} < \infty\}, \tag{3.17}$$

$$D(f_c) = \{y \in \mathbf{R}^n : \ \inf_{x \in D(f)} \{y'x - f(x)\} > -\infty\}. \tag{3.18}$$

Функция

$$f^c(y) = \sup_{x \in D(f)} \{ y'x - f(x) \}, \quad y \in D(f^c),$$

называется *сопряженной сверху (выпукло-сопряженной) функцией* для f и соответственно

$$f_c(y) = \inf_{x \in D(f)} \{ y'x - f(x) \}, \quad y \in D(f_c), \tag{3.19}$$

– сопряженной снизу (вогнуто-сопряженной) функцией для f.

Из этого определения непосредственно вытекает, что

$$(\operatorname{epi} f)^c = \operatorname{epi} f^c,$$

т.е. сопряженное сверху множество для верхнего полуцилиндра $\operatorname{epi} f$ совпадает с надграфиком функции f^c. Соответственно

$$(\operatorname{hypo} f)_c = \operatorname{hypo} f_c.$$

Между сопряженными снизу и сопряженными сверху функциями существует тесная взаимосвязь.

Т е о р е м а 3.5. *Для произвольной функции* f: $D(f) \to \mathbf{R}, D(f) \subseteq \mathbf{R}^n$, *справедливы соотношения*

$$(-f)_c(y) = -f^c(-y), \quad y \in D((-f)_c) = -D(f^c). \tag{3.20}$$

Д о к а з а т е л ь с т в о. Для $y \in D((-f)_c)$ имеем

$$(-f)_c(y) = \inf_{x \in D(-f) = D(f)} \{ y'x - (-f(x)) \} = \inf_{x \in D(f)} \{ -[(-y)'x - f(x)] \} =$$

$$= - \sup_{x \in D(f)} \{ (-y)'x - f(x) \} = -f^c(-y).$$

Из этой теоремы следует, что при изучении свойств сопряженных функций можно ограничиться в основном сопряженными сверху функциями. Поскольку в теории двойственности (см. § 5.2 и § 5.9) приходится рассматривать главным образом сопряженные сверху функции для выпуклых функций и лишь иногда – сопряженные снизу функции для вогнутых функций, то мы будем называть f^c просто сопряженной функцией.

В соответствии с геометрической интерпретацией множества M^c получается, что для множества

$$(y, w) \in (\operatorname{epi} f)^c = \operatorname{epi}(f^c), \quad \text{где } y \in D(f^c), \quad w = f^c(y),$$

гиперплоскость $H(y, f^c(y))$ представляет собой невертикальную опорную к $\operatorname{epi} f$ гиперплоскость. Обратно, если $H(y, w)$ – опорная к $\operatorname{epi} f$ гиперплоскость, то с необходимостью $y \in D(f^c)$ и $w = f^c(y)$. Если не существует таких гиперплоскостей, то

$$\sup_{x \in D(f)} \{ y'x - f(x) \} = \infty \quad \forall y \in \mathbf{R}^n,$$

и тем самым $D(f^c) = \phi$

Например, для функции $f(x) = x^3$, $x \in \mathbf{R}$, имеем

$$\sup_{x \in \mathbf{R}} \{ yx - x^3 \} = \infty \quad \forall y \in \mathbf{R}.$$

Пример 3.3. Пусть задана функция (рис. 3.7)

$$f(x) = \begin{cases} -x/2 & \text{при} \quad -2 \leqslant x < 0, \\ x(2-x) & \text{при} \quad 0 \leqslant x < 2, \\ x-2 & \text{при} \quad 2 \leqslant x < \infty. \end{cases}$$

С помощью приведенной выше геометрической интерпретации можно по заданному вектору нормали $(y, -1)'$ к гиперплоскости $H(y, w)$ последовательно построить сопряженную функцию f^c (рис. 3.8). А именно, получим

$$f^c(y) = \begin{cases} -2y-1 & \text{при} \quad -\infty < y < -1/2, \\ 0 & \text{при} \quad -1/2 \leqslant y < 0, \\ 2y & \text{при} \quad 0 \leqslant y \leqslant 1. \end{cases}$$

В частном случае $M = \operatorname{epi} f$. Из свойств, которые были установлены в § 3.1 для сопряженных множеств M^c, непосредственно вытекают следующие утверждения.

Теорема 3.6. 1) *Пусть заданы функции* $f_1\colon D(f_1) \to \mathbf{R}$, $f_2\colon D(f_2) \to \mathbf{R}$, *и* $D(f_1) \subset D(f_2)$. *Тогда, если существуют* f_1^c *и* f_2^c, *то*

$$f_1(x) \leqslant f_2(x) \quad \forall x \in D(f_1) \Rightarrow f_1^c(y) \geqslant f_2^c(y) \quad \forall y \in D(f_1^c). \tag{3.21}$$

2) *Если функция* $f\colon D(f) \to \mathbf{R}$ *такова, что существует* f^c, *то*

$$D(f^{cc}) \supseteq D(f); \quad f^{cc}(x) \leqslant f(x) \quad \forall x \in D(f).$$

3) *Если функция* $f\colon D(f) \to \mathbf{R}$ *такова, что существует* f^c, *то*

$$f^{ccc}(y) = f^c(y) \quad \forall y \in D(f^c) = D(f^{ccc}).$$

По теореме 2.77 для любой выпуклой функции $f\colon D(f) \to \mathbf{R}$ в каждой точке $x^0 \in \operatorname{ri} D(f)$ существует по меньшей мере одна невертикальная опорная гиперплоскость. Поэтому для выпуклых функций $D(f^c) \neq \phi$, и соответствующие предположения теоремы 3.6 для выпуклых функций могут быть опущены.

Из теоремы 3.1 сразу же получается следующая теорема.

Теорема 3.7. *Пусть* $f\colon D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$, – *произвольная функция, у которой* $D(f^c) \neq \phi$. *Тогда* f^c *на выпуклом множестве* $D(f^c)$ *является замкнутой снизу выпуклой функцией.*

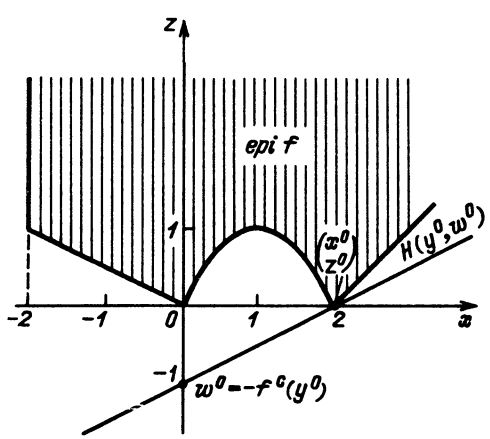


Рис. 3.7.

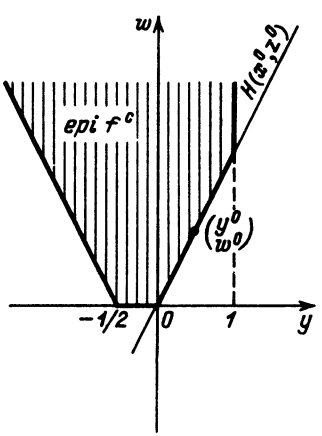


Рис. 3.8.

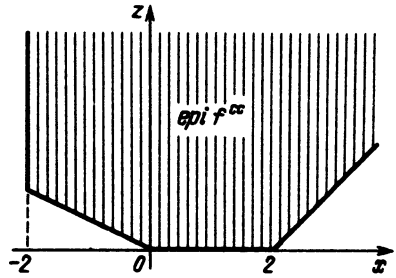


Рис. 3.9.

Соответственно из теоремы 3.2 вытекает следующая теорема.

Теорема 3.8. *Функция* f: $D(f) \to \mathbf{R}, D(f) \subseteq \mathbf{R}^n$, *тогда и только тогда замкнута снизу и выпукла, когда*

$$D(f^{cc}) = D(f); \qquad f^{cc}(x) = f(x) \qquad \forall\, x \in D(f).$$

Пример 3.4. Для приведенной в примере 3.3 функции f имеем (рис. 3.9)

$$f^{cc}(x) = \begin{cases} -x/2 & \text{при} \quad -2 \leqslant x < 0, \\ 0 & \text{при} \quad 0 \leqslant x < 2, \\ x-2 & \text{при} \quad 2 \leqslant x < \infty. \end{cases}$$

В следующем утверждении используется связь между сопряженными функциями для f и $\underline{f}$.

Теорема 3.9. *Для произвольной функции* f: $D(f) \to \mathbf{R}, D(f) \subseteq \mathbf{R}^n$, *у которой* $D(f^c) \neq \phi$, *имеем*

$$D((\underline{f})^c) = D(f^c), \quad (\underline{f})^c(y) = f^c(y) \qquad \forall\, y \in D(f^c).$$

Доказательство. Условие $D(f^c) \neq \phi$ обеспечивает существование f^c. Согласно теореме 3.7 f^c является замкнутой снизу выпуклой функцией на множестве $D(f^c)$. По теореме 2.84 $\operatorname{epi} f^c$ замкнуто тогда и только тогда, когда

$$f^c(y) = (\underline{f^c})(y) \qquad \forall\, y \in D(f^c).$$

Далее в силу теорем 3.6 и 3.8 имеем

$$(\underline{f})^c = (f^{cc})^c = (f^c)^{cc} = f^c = (\underline{f^c}).$$

Из теоремы 3.3 вытекает следующее утверждение о "наибольшей" (поточечно) замкнутой снизу выпуклой функции, которая лежит "ниже" заданной функции.

Теорема 3.10. *Пусть* f: $D(f) \to \mathbf{R}, D(f) \subseteq \mathbf{R}^n$, – *произвольная функция, у которой* $D(f^c) \neq \phi$. *Тогда* $g = f^{cc}$ – *"наибольшая" (поточечно) замкнутая снизу выпуклая функция, обладающая свойством*

$$f(x) \geqslant g(x) \qquad \forall\, x \in D(f).$$

Для дальнейшего изучения связи между f и f^{cc} приведем без доказательства следующую теорему.

Теорема 3.11. *Пусть* f: $D(f) \to \mathbf{R}, D(f) \subseteq \mathbf{R}^n$, – произвольная функция, $D(f^c) \neq \phi$ и $\widetilde{D} = \{x \in D(f) : \exists\, x^1, x^2 \in D(f), x \in (x^1, x^2), f(x) >$

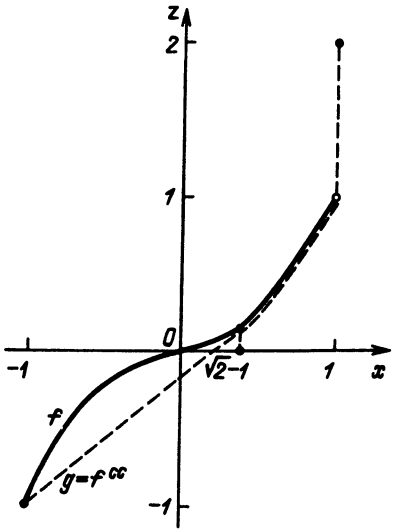


Рис. 3.10.

$> tf(x^1)+(1-t)f(x^2)\ \forall t\in(0,1)\}$. *Тогда*

1) $\mathbf{C}(D(f))\subseteq D(f^{cc})\subseteq\overline{\mathbf{C}(D(f))}$;

2) *ни через одну точку* $(x, f(x))'\in$ $\in\mathbf{R}^{n+1}$, *где* $x\in\widetilde{D}$, *не проходит гиперплоскость, опорная к* $\operatorname{epi} f$, *т.е.*

$$f(x)>f^{cc}(x)\quad\forall x\in\widetilde{D}.$$

Для некоторых функций f существуют выпуклые функции h: $D(f)\to$ $\to\mathbf{R}$, которые не являются замкнутыми снизу и для которых выполняется $h(x)\leqslant f(x)\ \forall x\in D(f)$, но $h(x^0)>$ $> f^{cc}(x^0)$ для некоторых граничных точек x^0 множества $D(f^{cc})$ и, в частности, для точек, принадлежащих $D(f)$.

Иллюстрацией к теореме 3.10 служит следующий пример.

Пример 3.5. Пусть функция f задана соотношениями $D(f)=[-1,1]\subset\mathbf{R}$ и

$$f(x)=\begin{cases}(\operatorname{sgn}x)\,x^2 & \text{при}\quad x\in[-1,1),\\ 2 & \text{при}\quad x=1.\end{cases}$$

С помощью простых выкладок получаем (см. рис. 3.10)

$$g(x)=f^{cc}(x)=\begin{cases}-1+2(\sqrt{2}-1)(x+1) & \text{при}\quad x\in[-1,\sqrt{2}-1],\\ x^2 & \text{при}\quad x\in[\sqrt{2}-1,1].\end{cases}$$

Функция h, получающаяся из g по правилу

$$h(x)=\begin{cases}g(x) & \text{при}\quad x\in[-1,1),\\ 2 & \text{при}\quad x=1,\end{cases}$$

дает пример ситуации, упомянутой после теоремы 3.11.

Из теоремы 4.4 вытекает интересная зависимость между субдифференциалами функций f и f^c.

Теорема 3.12. *Пусть* f: $D(f)\to\mathbf{R}, D(f)\subseteq\mathbf{R}^n$, – *произвольная функция и* $x^0\in D(f)$. *Справедливы соотношения:*

$$1)\ y^0\in\partial f(x^0)\Leftrightarrow f^c(y^0)+f(x^0)=y^{0\prime}x^0; \tag{3.22}$$

$$2)\ y^0\in\partial f(x^0)\Leftrightarrow x^0\in\partial f^c(y^0). \tag{3.23}$$

Доказательство. 1. В силу определений субдифференциала и функции f^c точка y^0 тогда и только тогда принадлежит субдифференциалу $\partial f(x^0)$, когда $H(y^0, f^c(y^0))$ является гиперплоскостью, опорной к $\operatorname{epi} f$, с опорной точкой $(x^0, f(x^0))'$, т.е. выполняется (3.22) и $\operatorname{epi} f\subseteq$ $\subset H_-(y^0, f^c(y^0))$.

2. Пусть $y^0\in\partial f(x^0)$. По доказанному выше $H(y^0, f^c(y^0))$ является гиперплоскостью, опорной к $\operatorname{epi} f$ с опорной точкой $(x^0, f(x^0))'$. Применим теперь теорему 3.4 к множеству $M=\operatorname{epi} f$. Она утверждает, что

$H(x^0, f(x^0))$ является гиперплоскостью, опорной к $(\operatorname{epi} f)^c = \operatorname{epi} f^c$ с опорной точкой $(y^0, f^c(y^0))'$. При этом необходимо, чтобы выполнялось $\operatorname{epi} f^c \subseteq H_-(x^0, f(x^0))$, что влечет $x^0 \in \partial f^c(y^0)$.

Точки x^0 и y^0, удовлетворяющие соотношению

$$f^c(y^0) + f(x^0) = y^{0\prime} x^0,$$

называются *сопряженными* (друг к другу) *точками.*

Если $f: D(f) \to \mathbf{R}$ – выпуклая функция, то по теореме 2.77 для каждого $x^0 \in \operatorname{ri} D(f)$ существует хотя бы одна сопряженная точка y^0.

Следующие три теоремы могут быть полезны при нахождении сопряженных функций.

Т е о р е м а 3.13. *Пусть* $f: D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$ – *произвольная функция, а* f^c – *сопряженная для* f. *Тогда функция, сопряженная для функции* $g(x) = \alpha f(x) + z'x + \beta$, *где* $x \in D(f)$, $z \in \mathbf{R}^n$, $\alpha, \beta \in \mathbf{R}$; $\alpha > 0$, *имеет вид* $g^c(y) = \alpha f^c((y - z)/\alpha)$, *где* $y \in D(g^c) = \alpha D(f^c) + z$.

Д о к а з а т е л ь с т в о. В силу равенства $D(f) = D(g)$ имеем

$$\alpha \sup_{x \in D(f)} \left\{ \left(\frac{y - z}{\alpha} \right)' x - f(x) \right\} - \beta = \sup_{x \in D(f)} \{(y - z)'x - \alpha f(x) - \beta\} =$$

$$= \sup_{x \in D(f)} \{ y'x - (\alpha f(x) + z'x + \beta) \} = \sup_{x \in D(g)} \{ y'x - g(x) \} .$$

Т е о р е м а 3.14. *Пусть* $f: \mathbf{R}^n \to \mathbf{R}$ – *произвольная функция,* A – *матрица размера* $m \times n$, $r(A) = m \leqslant n$, $b \in \mathbf{R}^n$. *Тогда*

$$f^c(Au - b) = h^c(u) \quad \forall u \in D(h^c), \tag{3.24}$$

где функция $h: \mathbf{R}^m \to \mathbf{R}$ *имеет вид*

$$h(z) = \inf_{y \in Y_z} \{ y'b + f(y) \}, \quad Y_z = \{ y \in \mathbf{R}^n : A'y = z \} .$$

Д о к а з а т е л ь с т в о. Для $u \in D(h^c)$ имеем

$$h^c(u) = \sup_{z \in \mathbf{R}^m} \{ u'z - h(z) \} =$$

$$= \sup_{z \in \mathbf{R}^m} \{ u'z + \sup_{y \in Y_z} \{ -y'b - f(y) \} \} = \sup_{z \in \mathbf{R}^m} \{ \sup_{y \in Y_z} \{ u'z - y'b - f(y) \} \} =$$

$$= \sup_{y \in \mathbf{R}^n} \{ (Au - b)'y - f(y) \} = f^c(Au - b).$$

Т е о р е м а 3.15. *Пусть* $f: D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$, – *сепарабельная функция вида*

$$f(x) = f(x^1, x^2) = f_1(x^1) + f_2(x^2), \quad (x^1, x^2)' \in D(f) = D(f_1) \times D(f_2),$$

$$D(f_1) \subseteq \mathbf{R}^k, \quad D(f_2) \subseteq \mathbf{R}^l, \quad k + l = n.$$

Тогда сопряженная функция f^c *также сепарабельна*:

$$f^c(y) = f^c(y^1, y^2) = f_1^c(y^1) + f_2^c(y^2),$$

$$(y^1, y^2)' \in D(f_1^c) \times D(f_2^c) \subseteq \mathbf{R}^k \times \mathbf{R}^l.$$

Доказательство. Для $y \in D(f^c)$ имеем

$$f^c(y) = \sup_{x \in D(f)} \{y'x - f(x)\} = \sup_{(x^1, x^2)' \in D(f)} \{y^{1'}x^1 + y^{2'}x^2 - f_1(x^1) - f_2(x^2)\} =$$

$$= \sup_{x^1 \in D(f_1^c)} \{y^{1'}x^1 - f_1(x^1)\} + \sup_{x^2 \in D(f_2^c)} \{y^{2'}x^2 - f_2(x^2)\} = f_1^c(y^1) + f_2^c(y^2).$$

Многократно применяя теорему 3.15 к функции

$$f(x) = f(x_1, \ldots, x_n) = \sum_{i=1}^{n} f_i(x_i), \quad x \in D(f),$$

получим

$$f^c(y) = \sum_{i=1}^{n} f_i^c(y_i) \quad \forall y = (y_1, \ldots, y_n)' \in D(f^c).$$

Завершим данный параграф теоремой о сопряженных для функций вида $f: \mathbf{R}^n \to \overline{\mathbf{R}}$ (см. п. 2.2.7).

Для $f: \mathbf{R}^n \to \overline{\mathbf{R}}$ сопряженная функция $f^c: \mathbf{R}^n \to \overline{\mathbf{R}}$ определяется согласно правилу

$$f^c(y) = \sup_{x \in \mathbf{R}^n} \{y'x - f(x)\} = \sup_{x \in \operatorname{dom} f} \{y'x - f(x)\}, \quad y \in \mathbf{R}^n. \tag{3.25}$$

Из теорем 3.2 и 2.89 вытекает следующая теорема.

Теорема 3.16. *Собственно выпуклая функция* $f: \mathbf{R}^n \to \overline{\mathbf{R}}$ *тогда и только тогда полунепрерывна снизу на* $\mathbf{R}^n$, *когда*

$$f(x) = f^{cc}(x) \quad \forall x \in \mathbf{R}^n.$$

Если функция $f: \mathbf{R}^n \to \overline{\mathbf{R}}$ *не является собственно выпуклой и* $f(x) = -\infty$ *хотя бы для одной точки* x, *то*

$$f^{cc}(x) = -\infty \quad \forall x \in \mathbf{R}^n.$$

3.3. Примеры сопряженных функций

Цель данного параграфа состоит в нахождении для некоторых важных классов функций соответствующих сопряженных (сверху) функций.

3.3.1. Аффинные функции. Для аффинных функций $f: D(f) \to \mathbf{R}$ вида

$$f(x) = a'x + b, \quad a \in \mathbf{R}^n, \quad b \in \mathbf{R}, \quad D(f) = \mathbf{R}^n, \tag{3.26}$$

имеем

$$\sup_{x \in \mathbf{R}^n} \{y'x - a'x - b\} = \sup_{x \in \mathbf{R}^n} \{(y - a)'x - b\} = \begin{cases} \infty & \text{при } y \neq a, \\ -b & \text{при } y = a, \end{cases}$$

и поэтому $D(f^c) = \{a\}$, $f^c(a) = -b$.

Функция (3.26) выпукла и замкнута снизу, следовательно, по теореме 3.8

$$f^{cc}(x) = f(x) \quad \forall x \in \mathbf{R}^n.$$

Теорема 3.17. *Аффинная функция* $f(x) = a'x + b$, $D(f) = \mathbf{R}^n$; $a \in \mathbf{R}^n$, $b \in \mathbf{R}$, *является сопряженной для функции* $g(y) = -b$, $D(g) = \{a\} \subset \mathbf{R}^n$; $b \in \mathbf{R}$.

3.2.2. Положительно однородные функции. Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ положительно однородна степени 1, т.е.

$$f(\lambda x) = \lambda f(x) \quad \forall x \in \mathbf{R}^n, \quad \forall \lambda \in \mathbf{R}_+.$$

В силу равенства

$$y'(\lambda x) - f(\lambda x) = \lambda(y'x - f(x)) \quad \forall \lambda \in \mathbf{R}_+$$

для произвольного фиксированного $x^0 \in \mathbf{R}^n$ получим

$$f^c(y) = \sup_{x \in \mathbf{R}^n} \{y'x - f(x)\} \geqslant \sup_{\lambda \in \mathbf{R}_+} \lambda \{y'x^0 - f(x^0)\}. \tag{3.27}$$

Если для заданного $y \in \mathbf{R}^n$ выполняется

$$y'x - f(x) \leqslant 0 \quad \forall x \in \mathbf{R}^n,$$

то

$$0 \geqslant f^c(y) = \sup_{x \in \mathbf{R}^n} \{y'x - f(x)\} \geqslant \sup_{\lambda \in \mathbf{R}_+} \lambda \{y'x^0 - f(x^0)\} = 0 \tag{3.28}$$

и тем самым $y \in D(f^c)$, $f^c(y) = 0$.

Если же для заданного $y \in \mathbf{R}^n$ существует единственный x^0, для которого $y'x^0 - f(x^0) > 0$, то из (3.27) следует

$$\sup_{x \in \mathbf{R}^n} \{y'x - f(x)\} = \infty, \quad y \notin D(f^c).$$

Поэтому справедливо следующее утверждение.

Теорема 3.18. *Пусть* $f\colon \mathbf{R}^n \to \mathbf{R}$ – *положительно однородная степени* 1 *выпуклая функция. Тогда* $D(f^c) = \{y \in \mathbf{R}^n\colon\ y'x - f(x) \leqslant 0 \quad \forall x \in \mathbf{R}^n\} \neq \phi$ *и* $f^c(y) = 0 \ \ \forall y \in D(f^c)$.

Пример 3.6. Евклидова норма в $\mathbf{R}^n$

$$f(x) = \|x\| = (x_1^2 + \ldots + x_n^2)^{1/2}, \quad x \in \mathbf{R}^n,$$

является положительно однородной функцией степени 1. Простые рассуждения показывают, что

$$D(f^c) = \{y \in \mathbf{R}^n\colon\ \|y\| \leqslant 1\}, \quad f^c(y) = 0 \quad \forall y \in D(f^c).$$

3.3.3. Кусочно-линейные функции. Для функции

$$f(x) = \max_{i=1,2} f_i(x), \quad x \in \mathbf{R}^n,$$

где

$$f_i(x) = a^{i\prime}x + b_i, \quad a^i \in \mathbf{R}^n, \quad b_i \in \mathbf{R}, \quad i = 1, 2; \quad a^1 \neq a^2,$$

имеем

$$D(f^c) = [a^1, a^2] = \{y\colon y = \lambda a^1 + (1 - \lambda)a^2, \ \lambda \in [0, 1]\},$$

$$f^c(y) = -\lambda b_1 - (1 - \lambda)b_2 \quad \forall y \in D(f^c).$$

Для доказательства этого утверждения необходимо рассмотреть два случая.

Случай 1. $y \notin [a^1, a^2]$. Множество $[a^1, a^2]$ компактно и выпукло. По второй теореме отделимости (теорема 2.36) существует гиперплоскость H, которая строго разделяет y и $[a^1, a^2]$. Следовательно, най-

дутся $x^0 \in \mathbf{R}^n$ и $\alpha \in \mathbf{R}$ такие, что

$$x^{0\prime} y = y' x^0 > \alpha = \max_{i=1,2} a^{i\prime} x^0. \tag{3.29}$$

Далее, справедливо равенство

$$\sup_{h>0} \{ y'(hx^0) - \max_{i=1,2} \{ a^{i\prime}(hx^0) + b_i \} \} =$$

$$= \sup_{h>0} h \{ y' x^0 - \max_{i=1,2} \{ a^{i\prime} x^0 + b_i / h \} \} .$$

В силу (3.29) полученное выражение не ограничено при $h \to \infty$. Поэтому $y \notin D(f^c)$.

Случай 2. $y \in [a^1, a^2]$. Тогда $y = \lambda a^1 + (1-\lambda) a^2$ для некоторого $\lambda \in [0,1]$ и потому

$$\sup_{x \in \mathbf{R}^n} \{ (\lambda a^1 + (1-\lambda) a^2)' x - f(x) \} = \sup_{x \in \mathbf{R}^n} \{ (\lambda a^1 + (1-\lambda) a^2)' x -$$

$$- \max_{i=1,2} \{ a^{i\prime} x + b_i \} \} \leqslant \lambda \sup_{x \in \mathbf{R}^n} \{ a^{1\prime} x - \max \{ a^{i\prime} x + b_i \} \} +$$

$$+ (1-\lambda) \sup_{x \in \mathbf{R}^n} \{ a^{2\prime} x - \max_{i=1,2} \{ a^{i\prime} x + b_i \} \} = \lambda(-b_1) + (1-\lambda)(-b_2).$$

В силу того, что $a^1 \neq a^2$, существует хотя бы один x^*, для которого $a^{1\prime} x^* + b_1 = a^{2\prime} x^* + b_2$. Для такого x^* имеем

$$(\lambda a^1 + (1-\lambda) a^2)' x^* - f(x^*) = (\lambda a^1 + (1-\lambda) a^2)' x^* - a^{1\prime} x^* - b_1 =$$

$$= \lambda a^{1\prime} x^* + (1-\lambda)(a^{1\prime} x^* + b_1 - b_2) - a^{1\prime} x^* - b_1 = \lambda(-b_1) + (1-\lambda)(-b_2).$$

Отсюда следует $y \in D(f^c)$ и $f^c(y) = \lambda(-b_1) + (1-\lambda)(-b_2)$.

3.3.4. Функции, постоянные на полупространстве. Пусть $a \in \mathbf{R}^n$, $a \neq 0$; $b, c \in \mathbf{R}$,

$$f(x) = b \quad \forall x \in D(f) = \{ x \in \mathbf{R}^n : a' x \leqslant c \}, \quad y \in \mathbf{R}^n.$$

Выделим следующие случаи.

Случай 1. Точка y не лежит в подпространстве, порожденном вектором a. Тогда для сколь угодно большого $d > 0$ имеем

$$\{ x : y' x > d \} \cap \{ x : a' x \leqslant c \} \neq \phi$$

и тем самым $\sup_{x \in D(f)} \{ y' x - f(x) \} = \infty$.

Случай 2. $y = 0$. При этом

$$f^c(0) = \sup_{x \in D(f)} \{ -f(x) \} = -b.$$

Случай 3. $y = \rho a$, $\rho > 0$. Для полупространства $H_-^g = \{ x : y' x \leqslant g \}$, $g \in \mathbf{R}$, выполняются включения

$$H_-^g \subseteq \{ x : a' x \leqslant c \} \quad \text{при } g/\rho \leqslant c,$$

$$H_-^g \supseteq \{ x : a' x \leqslant c \} \quad \text{при } g/\rho \geqslant c.$$

Полагая $\widetilde{H}_{-}^{g} = H_{-}^{g} \cap \{x: a'x \leqslant c\}$, получим

$$f^c(y) = f^c(\rho a) = \sup_{g \in \mathbf{R}} \{ \sup_{x \in \widetilde{H}_{-}^{g}} \{a'x - f(x)\} \} =$$

$$= \sup_{g \in \mathbf{R}} \{ \sup_{x \in \widetilde{H}_{-}^{g}} \{y'x - b\} \} = \rho c - b.$$

Случай 4. $y = \rho a$, $\rho < 0$. Поскольку при $g > \rho c$ полупространство H_{-}^{g} содержит все x, для которых $a'x = c$, то $\widetilde{H}_{-}^{g} \neq \phi$ для достаточно большого $g \in \mathbf{R}$. Поэтому

$$\sup_{x \in \mathbf{R}^n} \{y'x - f(x)\} = \sup_{g \in \mathbf{R}} \{ \sup_{x \in \widetilde{H}_{-}^{g}} \{y'x - f(x)\} \} = \sup_{g \in \mathbf{R}} \{g - b\} = \infty.$$

Подводя итог, получаем

$$D(f^c) = \{y: y = \rho a,\ \rho \geqslant 0\},\quad f^c(y) = \rho c - b \quad \forall y \in D(f^c).$$

3.3.5. Квадратичные формы. Пусть C – симметрическая матрица порядка $n \times n$, $d \in \mathbf{R}^n$, $Q(x) = \frac{1}{2}x'Cx$ – квадратичная форма, $f(x) = \frac{1}{2}x'Cx + d'x$, $x \in \mathbf{R}^n$. Если Q положительно полуопределена, то, очевидно, f – выпуклая функция.

Пусть M_N обозначает нуль-пространство матрицы C: $M_N = \{x: Cx = 0\}$, а M_R – пространство образов матрицы C: $M_R = \{y: Cx = y,\ x \in \mathbf{R}^n\}$. Множества M_N и M_R ортогональны друг к другу, т.е.

$$x'y = 0 \quad \forall (x, y) \in M_N \times M_R.$$

Как известно, $M_N \cap M_R = \{0\}$ и $M_N + M_R = \mathbf{R}^n$. Покажем, что $f^c(Cx) = \frac{1}{2}x'Cx$, $D(f^c) = M_R$.

Ограничимся вначале случаем $d = 0$.

Пусть $y \in D(f^c)$. Для произвольного $x^* \in M_N$ получим

$$\infty > f^c(y) = \sup_{x \in \mathbf{R}^n} \{y'x - f(x)\} \geqslant$$

$$\geqslant \sup_{\lambda \in \mathbf{R}} \{y'(\lambda x^*) - \frac{1}{2}(\lambda x^*)'C(\lambda x^*)\} = \sup_{\lambda \in \mathbf{R}} \lambda \{y'x^*\}.$$

Поэтому должно выполняться равенство $y'x^* = 0$, т.е. $y \in M_R$ и $D(f^c) \subseteq M_R$.

Чтобы показать справедливость противоположного включения, рассмотрим произвольный вектор $y \in M_R$. В этом случае $y^* = Cx^*, x^* \in \mathbf{R}^n$, и потому

$$f^c(y^*) = f^c(Cx^*) = \sup_{x \in \mathbf{R}^n} \left\{ x'Cx^* - \frac{1}{2}x'Cx \right\} =$$

$$= \frac{1}{2}x^{*\prime}Cx^* + \sup_{x \in \mathbf{R}^n} \left\{ -\frac{1}{2}(x - x^*)'C(x - x^*) \right\} = \frac{1}{2}x^{*\prime}Cx^* < \infty.$$

С помощью теоремы 3.13 получаем следующее утверждение.

Теорема 3.19. *Пусть* C – *симметрическая матрица порядка* $n \times n$, $Q(x) = \frac{1}{2}x'Cx$ – *квадратичная форма,* $f(x) = \frac{1}{2}x'Cx + d'x \quad \forall x \in \mathbf{R}^n$, $d \in \mathbf{R}^n$.

1) *Если* Q *положительно полуопределена, то*

$$D(f^c) = \{z = y + d:\ y \in M_R\}, \quad f^c(Cx + d) = \frac{1}{2}x'Cx \quad \forall x \in \mathbf{R}^n.$$

2) *Если* Q *положительно определена, то*

$$D(f^c) = \mathbf{R}^n, \quad f^c(y) = \frac{1}{2}(y - d)'C^{-1}(y - d).$$

Заметим что во втором случае в силу положительной определенности Q существует обратная матрица C^{-1} и уравнение $y = Cx + d$ имеет решение $x = C^{-1}(y - d)$.

Проиллюстрируем это замечание на примере.

Пример 3.7. Функция $f(x) = x^2$, $x \in \mathbf{R}$, является частным случаем квадратичной формы, у которой $C = (2)$ и $d = 0$. Следовательно, $f^c(y) = y^2/4$, $y \in D(f^c) = \mathbf{R}$.

3.4. Сопряженные для дифференцируемых функций

Пусть $f: \mathbf{R}^n \to \mathbf{R}$ – заданная выпуклая функция. Предположим, что для $y^0 \in \mathbf{R}^n$ существует $x^0 \in \mathbf{R}^n$ такое, что функция f дифференцируема в окрестности x^0 и выполняется равенство $y^0 = \nabla f(x^0)$. Тогда для $y^0 = \nabla f(x^0)$ имеем

$$f(x) \geqslant f(x^0) + \nabla f(x^0)'(x - x^0) = f(x^0) + y^{0\prime}(x - x^0) \quad \forall x \in \mathbf{R}^n.$$

Это влечет неравенство

$$y^{0\prime}x^0 - f(x^0) \geqslant y^{0\prime}x - f(x) \quad \forall x \in \mathbf{R}^n.$$

Отсюда, в свою очередь, следует

$$y^0 \in D(f^c), \quad f^c(y^0) = y^{0\prime}x^0 - f(x^0).$$

Таким образом, $y^0 = \nabla f(x^0)$ и x^0 – сопряженные точки. Касательная гиперплоскость $H(y^0, f^c(y^0))$, проходящая через точку $(x^0, f(x^0))'$, является, очевидно, опорной к epi$f$.

Пусть теперь $f: \mathbf{R}^n \to \mathbf{R}$ – дифференцируемая функция, для которой $y = g(x) = \nabla f(x)$, $x \in \mathbf{R}^n$, является взаимно однозначным отображением пространства $\mathbf{R}^n$ на себя с обратным отображением $x = \varphi(y)$. Тогда функция f^c может быть определена явно:

$$f^c(y) = y'\varphi(y) - f(\varphi(y)), \quad y \in \mathbf{R}^n. \tag{3.30}$$

Если не налагать на f достаточно сильных предположений, то нельзя ожидать, что приведенные выше рассуждения дадут явное представление функции f^c. Дело в том, что множество образов g будет, вообще говоря, собственным подмножеством пространства $\mathbf{R}^n$. Кроме того, не каждый $y \in D(f^c)$ можно представить в форме $y = \nabla f(x)$ при подходящем x, даже если f дифференцируема для всех $x \in \operatorname{int} D(f)$. Чтобы обойти трудности, которые могли бы встретиться по этой причине при выводе теорем

двойственности, мы ограничимся далее такими функциями, которые удовлетворяют определенным дополнительным "условиям регулярности" (см. в этой связи, например, работу Стоера и Витцгала [79]).

Мы сформулируем следующее условие регулярности (R). Пусть $f: D(f) \to \mathbf{R}$, $D(f) \subseteq \mathbf{R}^n$, – выпуклая функция такая, что через каждую относительную граничную точку множества $\operatorname{epi} f$ проходит самое большее одна невертикальная опорная гиперплоскость. Функция f удовлетворяет условию регулярности (R) тогда и только тогда, когда:

а) $\operatorname{int} D(f) \neq \phi$;

б) f дифференцируема на $\operatorname{int} D(f)$ (отсюда в силу теоремы 2.82 через каждую точку $(x, f(x))'$, $x \in \operatorname{int} D(f)$, проходит в точности одна невертикальная опорная гиперплоскость);

в) все остальные гиперплоскости, опорные к $\operatorname{epi} f$ (т.е. исключая опорные гиперплоскости, проходящие через точки $(x, f(x))'$, $x \in \operatorname{int} D(f)$), являются вертикальными.

Условие регулярности (R) – весьма сильное требование, накладываемое на функцию f. Например, оно не выполняется для следующей замкнутой снизу выпуклой функции:

$$f(x) = x^2, \quad x \in D(f) = [-1, 1], \quad D(f) \subseteq \mathbf{R}.$$

Действительно, уравнение $yx - z = y - 1$ для каждого $y \geqslant 2$ определяет невертикальную гиперплоскость, опорную к $\operatorname{epi} f$ в точке $(1, 1)' \in \mathbf{R}^2$.

Имеет место следующая теорема.

Теорема 3.20. *Пусть $D(f) \subseteq \mathbf{R}^n$ – открытое выпуклое множество и $f: D(f) \to \mathbf{R}$ – выпуклая функция, удовлетворяющая условию регулярности (R). Для любого $y^0 \in \operatorname{ri} D(f^c)$, существует $x^0 \in \operatorname{int} D(f)$ такой, что $y^0 = \nabla f(x^0)$ и*

$$f^c(y^0) = f^c(\nabla f(x^0)) = [\nabla f(x^0)]' x^0 - f(x^0).$$

Доказательство. Теорема 3.12 в применении к f^c гарантирует для $y^0 \in \operatorname{ri} D(f^c)$ существование невертикальной гиперплоскости $H = \{(y, w)': x^0 y - w = r^0\}$, опорной к $\operatorname{epi} f^c$ в точке $(y^0, w^0)'$, где $w^0 = f^c(y^0)$. При этом $r^0 = f(x^0)$, а $\widetilde{H} = \{(x, z)': y^0 x - z = w^0\}$ – невертикальная гиперплоскость, опорная к $\operatorname{epi} f$ в точке $(x^0, f(x^0))'$. Согласно замечанию к условию (R) отсюда вытекает $x^0 \in \operatorname{int} D(f)$.

Для дифференцируемой замкнутой снизу выпуклой на открытом множестве $D(f)$ функции f условие регулярности (R) всегда выполнено. Поэтому для таких функций справедливо утверждение теоремы 3.20.

Пример 3.8. Функция

$$f(x) = \begin{cases} (x-1)^3/3 & \text{при } x \in (1, \infty), \\ 0 & \text{при } x \in [-1, 1], \ D(f) = \mathbf{R}, \\ -(x+1)^3/3 & \text{при } x \in (-\infty, -1) \end{cases}$$

замкнута снизу, выпукла и дифференцируема. Ее производная имеет вид

$$f'(x) = \begin{cases} (x-1)^2 & \text{при } x \in (1, \infty), \\ 0 & \text{при } x \in [-1, 1], \\ -(x+1)^2 & \text{при } x \in (-\infty, -1) \end{cases}$$

и поэтому $D(f^C) = \mathbf{R}$. Для каждого $y \in \mathbf{R}$ по теореме 3.20 найдется хотя бы один $x \in \mathbf{R}$ такой, что $f'(x) = y$. При этом для всех $x \in [-1, 1]$ выполняется равенство $y = 0 = f'(x)$, следовательно, $y = 0$ – сопряженная точка для каждого $x \in [-1, 1]$. В случае $y < 0$ соответствующая сопряженная точка x получается однозначно путем решения уравнения $y = -(x+1)^2$ относительно x: $x = -(-y)^{1/2} - 1$. В случае $y > 0$ имеем соответственно $x = y^{1/2} + 1$

Согласно теореме 3.20

$$f^C(y) = \begin{cases} y + 2y^{3/2}/3 & \text{при } y > 0, \\ 0 & \text{при } y = 0, \\ -[y - 2(-y)^{3/2}/3] & \text{при } y < 0. \end{cases}$$

Функция f^C не дифференцируема в точке $y = 0$ и не удовлетворяет также условию регулярности (R).

Пример 3.9. Функция

$$f(x) = \begin{cases} -\sqrt{x^2 - 1} & \text{при } x \in [\sqrt{5}, \infty), \\ -\sqrt{5}x/2 + 1/2 & \text{при } x \in (-\infty, \sqrt{5}), \end{cases} \quad D(f) = \mathbf{R},$$

выпукла и дифференцируема. Ее производная имеет вид

$$f'(x) = \begin{cases} -\dfrac{x}{\sqrt{x^2 - 1}} & \text{при } x \in [\sqrt{5}, \infty), \\ -\sqrt{5}/2 & \text{при } x \in (-\infty, \sqrt{5}), \end{cases}$$

и поэтому

$$-\sqrt{5}/2 \leqslant f'(x) < -1 \quad \forall x \in \mathbf{R}. \tag{3.31}$$

Справедливы соотношения:

$$f^C(-1) = \max\left\{ \sup_{x \geqslant \sqrt{5}} (-x + \sqrt{x^2 - 1}), \sup_{x < \sqrt{5}} \left| -x + \frac{\sqrt{5}}{2}x - \frac{1}{2} \right| \right\} =$$

$$= \lim_{x \to \infty} (-x + \sqrt{x^2 - 1}) = 0;$$

$$\sup_{x \in \mathbf{R}} \{yx - f(x)\} = \infty \quad \text{для } y > -1 \text{ и для } y < -\sqrt{5}/2.$$

Таким образом, с учетом (3.31) имеем $D(f^C) = [-\sqrt{5}/2, -1]$.

Для любого $y^0 \in [-\sqrt{5}/2, -1)$ найдется $x^0 \in \mathbf{R}$ такой, что

$$f^C(y^0) = y^0 x^0 - f(x^0). \tag{3.32}$$

Для $y^0 = -\sqrt{5}/2$ равенство (3.32) выполняется даже при всех $x < \sqrt{5}$. С помощью подстановки $y = f'(x)$ получаем следующее явное представление функции f^C:

$$f^C(x) = y\sqrt{\frac{y^2}{y^2 - 1}} + \sqrt{\frac{1}{y^2 - 1}} = -\sqrt{y^2 - 1}, \qquad y \in [-\sqrt{5}/2, -1). \tag{3.33}$$

При этом следует обратить внимание на то, что хотя точка $y^0 = -1$ и принадлежит множеству $D(f^C)$, тем не менее для y^0 не существует сопряженной точки $x^0 \in D(f)$.

Пример 3.10. Функция

$$f(x) = f(x_1, x_2) = x_2^2 - \sqrt{x_1^2 - 1},$$

$(x_1, x_2)' \in D(f) = \{(x_1, x_2)' \in \mathbf{R}^2 : x_1 \geqslant 1, x_2 \geqslant 0\}$, является сепарабельной: $f(x) = f_1(x_1) + f_2(x_2)$, где $f_1(x_1) = -\sqrt{x_1^2 - 1}$, $f_2(x_2) = x_2^2$.

По теореме 3.15 имеем

$$f^c(y) = f_1^c(y_1) + f_2^c(y_2), \quad y = (y_1, y_2)' \in D(f_1^c) \times D(f_2^c).$$

Сначала мы определим f_2^c:

$$f_2^c(y_2) = \sup_{x_2 \in \mathbf{R}_+} \{y_2 x_2 - f_2(x_2)\}.$$

Так как $y_2 x_2 - x_2^2 \leqslant 0$ для любых $y_2 \leqslant 0$ и $x_2 \geqslant 0$, то

$$f_2^c(y_2) = \sup_{x_2 \geqslant 0} \{y_2 x_2 - x_2^2\} = 0, \quad y_2 \leqslant 0.$$

Применяя теорему 3.20, определим $f_2^c(y_2)$ для $y_2 > 0$. Получим

$$D(f_2^c) = \mathbf{R}; \qquad f_2^c(y_2) = \begin{cases} y_2^2/4 & \text{при } y_2 \in \mathbf{R}_+, \\ 0 & \text{при } y_2 < 0. \end{cases}$$

Теперь определим f_1^c. Для произвольного $y_1 \leqslant -1$ выполняется неравенство $y_1 x_1 + \sqrt{x_1^2 - 1} \leqslant 0$ при всех $x_1 \geqslant 0$. Для $y_1 = -1$ имеем

$$\lim_{x_1 \to \infty} (y_1 x_1 + \sqrt{x_1^2 - 1}) = \lim_{x_1 \to \infty} \frac{-1}{\sqrt{x_1^2 - 1} + x_1} = f^c(-1) = 0.$$

Для произвольного фиксированного $y_1 < -1$ известными методами дифференциального исчисления получаем

$$f_1^c(y_1) = y_1 x_1 + \sqrt{x_1^2 - 1} = -\sqrt{y_1^2 - 1}.$$

Для произвольного фиксированного $y_1 > -1$ имеем

$$\lim_{x_1 \to \infty} (y_1 x_1 + \sqrt{x_1^2 - 1}) = \lim_{x_1 \to \infty} x_1 \left(y_1 + \sqrt{1 - \frac{1}{x_1^2}} \right) = \infty.$$

Поэтому $D(f_1^c) = (-\infty, -1]$.

Подводя итог, заключаем:

$$D(f^c) = (-\infty, -1] \times \mathbf{R},$$

$$f^c(y_1, y_2) = \begin{cases} y_2^2/4 - \sqrt{y_1^2 - 1} & \text{при } y_1 \leqslant -1, \quad y_2 > 0, \\ -\sqrt{y_1^2 - 1} & \text{при } y_1 \leqslant -1, \quad y_2 \leqslant 0. \end{cases}$$

Пример 3.11. Пусть $f(x) = f(x_1, \ldots, x_n) = \frac{1}{p} \sum_{i=1}^{n} |x_i|^p$, $x \in \mathbf{R}^n$; $p > 1$. Для произвольного $y \in \mathbf{R}^n$ имеем

$$\sup_{x \in \mathbf{R}^n} \left\{ y'x - \frac{1}{p} \sum_{i=1}^{n} |x_i|^p \right\} = \sup_{x \in \mathbf{R}^n} \left\{ \sum_{i=1}^{n} y_i x_i - \frac{1}{p} \sum_{i=1}^{n} |x_i|^p \right\} \geqslant 0.$$

Поскольку $\sup_{x \in \mathbf{R}^n} \left\{ y'x - \frac{1}{p} \sum_{i=1}^{n} |x_i|^p \right\} \leqslant \sup_{x \in \mathbf{R}^n} \left\{ \sum_{i=1}^{n} \left(|y_i||x_i| - \frac{1}{p} |x_i|^p \right) \right\}$

и $\left\{ \sum_{i=1}^{n} \left(|y_i||x_i| - \frac{1}{p} |x_i|^p \right) \right\} \to -\infty$ при $\|x\| \to +\infty$, то $\sup_{x \in \mathbf{R}^n} \left\{ y'x - \frac{1}{p} \sum_{i=1}^{n} |x_i|^p \right\}$

достигается по крайней мере в одной точке $x^* \in \mathbf{R}^n$ с $\| x^* \| < +\infty$. Функция $\left\{ y'x - \right.$

$- \left. \frac{1}{p} \sum_{i=1}^{n} |x_i|^p \right\}$, $x \in \mathbf{R}^n$, $p > 1$, дифференцируема для произвольного фиксированного $y \in \mathbf{R}^n$. Поэтому

$$\frac{\partial}{\partial x_i} \left\{ y'x - \frac{1}{p} \sum_{i=1}^{n} |x_i|^p \right\} \Bigg|_{x = x^*} = y_i - \operatorname{sgn} x_i^* \, |x_i^*|^{p-1} = 0, \quad i = 1, \ldots, n.$$

Отсюда следует $\operatorname{sgn} x_i^* = \operatorname{sgn} y_i$ и $|x_i^*| = |y_i|^{1/(p-1)}$, $i = 1, \ldots n$.

Таким образом, для заданного $y \in \mathbf{R}^n$ имеется ровно один $x \in \mathbf{R}^n$, для которого выполнены необходимые условия оптимальности

$$\frac{\partial}{\partial x_i} \left\{ y'x - \frac{1}{p} \sum_{i=1}^{n} |x_i|^p \right\} = 0, \qquad i = 1, \ldots, n.$$

Искомый супремум должен достигаться в этой точке. Поэтому получаем

$$D(f^c) = \mathbf{R}^n,$$

$$f^c(y) = \frac{p-1}{p} \sum_{i=1}^{n} |x_i^*|^p = \frac{1}{q} \sum_{i=1}^{n} |y_i|^q; \quad q = \frac{p}{p-1}.$$

4.1. Постановка задачи

Критериями оптимальности называют необходимые и (или) достаточные условия для того, чтобы вектор $x^0 \in \mathbf{R}^n$ являлся решением задачи математического программирования. Критерии оптимальности важны как при теоретических исследованиях, так и при численном решении задач. Ниже мы выведем *локальные* и *глобальные критерии оптимальности* (в том числе так называемые *критерии седловой точки*).

С этой целью мы рассмотрим общую задачу математического программирования

$$P\colon \min\{f(x)\colon\ x \in G \subseteq \mathbf{R}^n\},$$

где $f\colon \mathbf{R}^n \to \mathbf{R}$, а также ее частные случаи – задачи P_1 и P_1^*, допустимые области которых определяются с помощью неравенств и (или) уравнений (см. (1.4)). Задача P_1 имеет вид

$$P_1\colon \min\{f(x)\colon\ x \in G\},$$

где $G = \{x \in \mathbf{R}^n\colon\ g(x) \leqslant 0\}$, $f\colon\ \mathbf{R}^n \to \mathbf{R}$, $g\colon \mathbf{R}^n \to \mathbf{R}^m$.

Обозначая $I = \{1, \ldots, m\}$, $g = (g_1, \ldots, g_m)'$, допустимое множество G задачи P_1 можем представить в форме

$$G = \{x \in \mathbf{R}^n\colon\ g_i(x) \leqslant 0,\ i \in I\}.$$

Задача P_1^* записывается следующим образом:

$$P_1^*\colon \min\{f(x)\colon\ x \in G^*\},$$

где $G^* = \{x \in \mathbf{R}^n\colon\ g(x) \leqslant 0, h(x) = 0\}$, $f\colon\ \mathbf{R}^n \to \mathbf{R}$, $g\colon\ \mathbf{R}^n \to \mathbf{R}^m$, $h\colon\ \mathbf{R}^n \to \mathbf{R}^p$; $m, p \geqslant 1$. Множество G^* можно представить также в виде

$$G^* = \{x \in \mathbf{R}^n\colon\ g_i(x) \leqslant 0,\ i \in I;\ h_j(x) = 0,\ j \in I^*\},$$

где $g = (g_1, \ldots, g_m)'$, $h = (h_1, \ldots, h_p)'$, $I = \{1, \ldots, m\}$, $I^* = \{1, \ldots, p\}$.

В 4.2.1 приводятся локальные критерии оптимальности для задачи P (см. также [39]), в то время как в п. 4.2.2 и 4.2.4 развивается локальная теория множителей Лагранжа для задач P_1 и P_1^* соответственно (см., например, [27, 56, 59, 71, 96]).

Классическая теория множителей Лагранжа для задач с ограничениями в форме равенств уже излагалась сокращенно в § 1.3. Обобщая рассмотренные выше вопросы, мы обсудим в § 4.3 глобальную теорию множителей Лагранжа для задачи P_1, при этом существенным требованием окажется выпуклость функций f и g.

4.2. Локальные критерии оптимальности

4.2.1. Локальные критерии оптимальности для задачи P. Пусть в дальнейшем f: $\mathbf{R}^n \to \mathbf{R}$ – дифференцируемая функция. На множество $G \subseteq \mathbf{R}^n$ не налагается никаких специальных предположений.

Вначале мы рассмотрим сходящиеся последовательности $\{x^k\} \subseteq \mathbf{R}^n$ специального вида. Последовательность $\{x^k\}$ такая, что $x^k \to x^0$ $(k \to \infty)$, называется *сходящейся к x^0 в направлении $z \in \mathbf{R}^n$* (кратко: *направленно сходящейся к x^0*), если

$$\lim_{k \to \infty} \frac{x^k - x^0}{\| x^k - x^0 \|} = z, \quad \| z \| = 1. \tag{4.1}$$

При выполнении условия (4.1) мы будем писать

$$x^k \overset{z}{\to} x^0. \tag{4.2}$$

Очевидно, из любой сходящейся к x^0 последовательности $\{x^k\}$ можно выбрать направленно сходящуюся подпоследовательность, если $x^k \neq x^0$ для бесконечно большого числа номеров k. В самом деле, если $\{x^k\}$ – сходящаяся к x^0 последовательность и $x^k \neq x^0$ для $k \geqslant k_0$, то все члены последовательности $\{z^k\}$, где

$$z^k = \frac{x^k - x^0}{\| x^k - x^0 \|}, \qquad k \geqslant k_0,$$

являются элементами компактного множества $\{ x \in \mathbf{R}^n: \ \| x \| = 1\}$. Поэтому $\{z^k\}$ имеет хотя бы одну предельную точку z. Следовательно, в $\{z^k\}$ существует подпоследовательность, сходящаяся к z, а тем самым и направленно сходящаяся к x^0 подпоследовательность из $\{x^k\}$.

Следующее утверждение о дифференцируемых функциях можно считать некоторым обращением свойств производной по направлению.

Т е о р е м а 4.1. *Пусть последовательность $\{x^k\}$ сходится к x^0 в направлении z. Тогда*:

1) *если функция f: $\mathbf{R}^n \to \mathbf{R}$ дифференцируема в точке x^0, то*

$$\lim_{k \to \infty} \frac{f(x^k) - f(x^0)}{\| x^k - x^0 \|} = z' \nabla f(x^0); \tag{4.3}$$

2) *если функция f: $\mathbf{R}^n \to \mathbf{R}$ дважды дифференцируема в точке x^0, то*

$$\lim_{k \to \infty} \frac{f(x^k) - f(x^0) - (x^k - x^0)' \nabla f(x^0)}{\| x^k - x^0 \|^2} = \frac{1}{2} z' \nabla^2 f(x^0) z. \tag{4.4}$$

Д о к а з а т е л ь с т в о. 1. По предположению

$$\lim_{x \to x^0} \frac{f(x) - f(x^0) - (x - x^0)' \nabla f(x^0)}{\| x - x^0 \|} = 0.$$

Кроме того, существует предел

$$\lim_{x^k \overset{z}{\to} x^0} \frac{(x^k - x^0)'}{\| x^k - x^0 \|} \nabla f(x^0) = z' \nabla f(x^0).$$

Отсюда

$$\lim_{x^k \xrightarrow{z} x^0} \frac{f(x^k) - f(x^0) - (x^k - x^0)'\nabla f(x^0)}{\| x^k - x^0 \|} + \lim_{x^k \xrightarrow{z} x^0} \frac{(x^k - x^0)'}{\| x^k - x^0 \|} \nabla f(x^0) =$$

$$= \lim_{x^k \xrightarrow{z} x^0} \frac{f(x^k) - f(x^0)}{\| x^k - x^0 \|} = z'\nabla f(x^0).$$

2. Доказательство второго утверждения проводится аналогично.

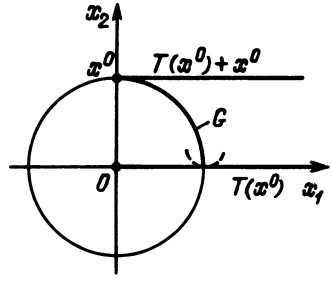


Рис. 4.1.

Как обобщение понятия касательной гиперплоскости к гиперповерхности, мы определим *касательный конус* $T(x^0)$ к множеству G в точке $x^0 \in G$:

$$T(x^0) = \{ \lambda z: \ \| z \| = 1, \ \exists \{ x^k \} \subset G, \ x^k \xrightarrow{z} x^0; \ \lambda \in \mathbf{R}_+ \}. \qquad (4.5)$$

Для любой внутренней точки x^0 множества G, очевидно, справедливо $T(x^0) = \mathbf{R}^n$. Для изолированных точек x^0 мы полагаем $T(x^0) = \{0\}$.

Как легко видеть, множество $T(x^0)$ действительно является конусом.

Пример 4.1. Пусть $G = \{ x \in \mathbf{R}^2: \ x_1^2 + x_2^2 = 1, x_1 \geqslant 0, \ 0 < x_2 < 2 \}$. Для точки $x^0 = (0, 1)'$ имеем (рис. 4.1)

$T(x^0) = \{ z \in \mathbf{R}^2: \ z_2 = 0; \ z_1 \geqslant 0 \}$.

В следующей теореме сформулировано важное свойство касательного конуса.

Теорема 4.2. *Для любого $x^0 \in G$ касательный конус $T(x^0)$ замкнут.*

Доказательство. Достаточно показать, что для любой последовательности $\{ z^l \}$ такой, что $z^l \to z$, $z^l \in T(x^0)$, $\| z^l \| = 1$, справедливо $z \in T(x^0)$. По предположению для каждого z^l существует последовательность $\{ x^{k,l} \}$, для которой $x^{k,l} \xrightarrow{z^l} x^0$. Поэтому для любого натурального числа s можно найти номера l_s и k_s такие, что

$$\| x^{k_s, l_s} - x^0 \| < \frac{1}{s}, \quad \left\| \frac{x^{k_s, l_s} - x^0}{\| x^{k_s, l_s} - x^0 \|} - z^{l_s} \right\| < \frac{1}{2s},$$

$$\| z^{l_s} - z \| < \frac{1}{2s}.$$

Следовательно,

$$\left\| \frac{x^{k_s, l_s} - x^0}{\| x^{k_s, l_s} - x^0 \|} - z \right\| \leqslant \left\| \frac{x^{k_s, l_s} - x^0}{\| x^{k_s, l_s} - x^0 \|} - z^{l_s} \right\| +$$

$$+ \| z^{l_s} - z \| < \frac{1}{2s} + \frac{1}{2s} = \frac{1}{s}.$$

Тем самым последовательность $\{x^s\} \subset G$, $x^s = x^{k_s, l_s}$, сходится к x^0 в направлении z, т.е. $z \in T(x^0)$.

Полученные нами предварительные результаты позволяют сформулировать необходимые критерии для существования локального минимума функции f на множестве G.

Теорема 4.3. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *дифференцируема в точке* $x^0 \in G$. *Если* f *достигает в* $x^0 \in G$ *локального минимума на* G, *то*

$$\nabla f(x^0)' z \geqslant 0 \quad \forall z \in T(x^0).$$

Доказательство. Для $z = 0$ имеем $\nabla f(x^0)' z = 0$. Пусть $z \in T(x^0)$, $z \neq 0$. Тогда существуют $\lambda \in \mathbf{R}_+$ и $z^* \in T(x^0)$, $\| z^* \| = 1$, такие, что $z = \lambda z^*$. Пусть, далее, $\{x^k\} \subset G$ – последовательность, для которой $x^k \xrightarrow{z^*} x^0$.

Так как функция f достигает в x^0 локального минимума на G, то для достаточно большого k выполняется неравенство $f(x^k) \geqslant f(x^0)$. Отсюда с учетом теоремы 4.1 следует

$$\lim_{k \to \infty} \frac{f(x^k) - f(x^0)}{\| x^k - x^0 \|} = \nabla f(x^0)' z^* \geqslant 0,$$

а потому и $\nabla f(x^0)' z \geqslant 0$.

Теорема 4.4. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *дважды дифференцируема в точке* $x^0 \in G$. *Если* f *достигает в* x^0 *локального минимума на* G *и* $\nabla f(x^0) = 0$, *то*

$$z' \nabla^2 f(x^0) z \geqslant 0 \quad \forall z \in T(x^0)\ ^{*)}.$$

Доказательство. Пусть $\{x^k\} \subset G$ – последовательность такая, что $x^k \xrightarrow{z} x^0$. В силу соотношений $\nabla f(x^0)'(x^k - x^0) = 0$ и $f(x^k) - f(x^0) \geqslant 0$ с учетом теоремы 4.1 для достаточно большого k имеем

$$\lim_{k \to \infty} \frac{f(x^k) - f(x^0) - \nabla f(x^0)'(x^k - x^0)}{\| x^k - x^0 \|^2} = \frac{1}{2} z' \nabla^2 f(x^0) z \geqslant 0.$$

В следующей теореме сформулировано достаточное условие для существования собственного локального минимума функции f на множестве G.

Теорема 4.5. *Пусть функция* $f\colon \mathbf{R}^n \to \mathbf{R}$ *дважды дифференцируема в точке* $x^0 \in G$. *Если* $\nabla f(x^0) = 0$, $z' \nabla^2 f(x^0) z > 0$ $\forall z \in T(x^0)$, $z \neq 0$, *то* f *достигает в* x^0 *собственного локального минимума на* G.

---

*) Следует особо указать на то, что функция f может достигать в x^0 локального минимума на G и без выполнения условия $\nabla f(x^0) = 0$.

Доказательство. Проведем доказательство от противного и предположим, что f не достигает в x^0 собственного локального минимума на G. Тогда существует последовательность $\{x^k\} \subset G$ такая, что $x^k \to x^0$ и $f(x^k) \leqslant f(x^0)$ $\forall k \in \mathbf{N}$. Без ограничения общности мы можем также считать, что $x^k \overset{\tilde{z}}{\to} x^0$. Следовательно, согласно теореме 4.1

$$\lim_{k \to \infty} \frac{f(x^k) - f(x^0)}{\| x^k - x^0 \|^2} = \frac{1}{2} \tilde{z}' \nabla^2 f(x^0) \tilde{z} \leqslant 0, \quad \tilde{z} \in T(x^0),$$

что противоречит условиям доказываемой теоремы.

Замечание. Аналогично теореме 4.5 доказывается следующее достаточное условие оптимальности.

Пусть функция f: $\mathbf{R}^n \to \mathbf{R}$ дифференцируема в точке $x^0 \in G$. Если $\nabla f(x^0)' z > 0$ $\forall z \in T(x^0), z \neq 0$, то f достигает в x^0 собственного локального минимума на G.

При $x^0 \in \operatorname{int} G$ имеем $T(x^0) = \mathbf{R}^n$, и из теорем 4.3 – 4.5 вытекают известные необходимые и достаточные условия для существования локального минимума функции f: $\mathbf{R}^n \to \mathbf{R}$.

По аналогии с предыдущими теоремами легко можно сформулировать утверждения о существовании локального максимума функции f на множестве G или на всем $\mathbf{R}^n$.

4.2.2. **Локальная теория множителей Лагранжа.** Применяя результаты 4.2.1, выведем локальные критерии оптимальности для задачи P_1. В дальнейшем g: $\mathbf{R}^n \to \mathbf{R}^m$ – дифференцируемая функция.

Неравенство $g_i(x) \leqslant 0$, $i \in I$, называется *активным ограничением* в точке $x^0 \in G$, если $g_i(x^0) = 0$.

Для $x^0 \in G$ положим $I_0 = \{i \in I: g_i(x^0) = 0\}$. Пусть вначале $I_0 \neq \phi$. Фучкции g_i, $i \in I_0$, соответствующие активным в точке $x^0 \in G$ ограничениям, заменим на аффинные функции

$$\nabla g_i(x^0)'(x - x^0), \quad x \in \mathbf{R}^n; \qquad i \in I_0. \tag{4.6}$$

Если из векторов $\nabla g_i(x^0), i \in I_0$, составить матрицу $\nabla g_{I_0}(x^0)$, то вместо (4.6) получим

$$\nabla g_{I_0}(x^0)'(x - x^0). \tag{4.7}$$

Обозначим через $|I_0|$ число элементов множества I_0. Тогда $\nabla g_{I_0}(x^0)$ – матрица порядка $n \times |I_0|$. Соответственно будем обозначать в дальнейшем $|I_0|$-мерный вектор, координаты которого определяются индексами активных ограничений (при сохранении их порядка следования), через y_{I_0}. Для наглядности множество

$$\{x: \nabla g_{I_0}(x^0)'(x - x^0) \leqslant 0, \; g_{I \setminus I_0}(x^0) + \nabla g_{I \setminus I_0}(x^0)'(x - x^0) \leqslant 0\}$$

можно было бы рассматривать как "линейную аппроксимацию" G в точке x^0. Однако, как мы это еще увидим ниже (см. пример 4.2), подобная интерпретация в некоторых случаях проблематична.

Множество

$$K(x^0)=\{z\in\mathbf{R}^n:\ \nabla g_{I_0}(x^0)'z\leqslant 0\} \tag{4.8}$$

назовем *линеаризующим конусом* для G в точке $x^0\in G$. Если $I_0=\phi$, мы полагаем $K(x^0)=\mathbf{R}^n$.

Изучим теперь связь между конусами $K(x^0)$ и $T(x^0)$.

Теорема 4.6. *Справедливо включение* $T(x^0)\subseteq K(x^0)$.

Доказательство. Для $I_0=\phi$ утверждение теоремы тривиально, поскольку $I_0=\phi$ влечет $x^0\in\operatorname{int}G$. Поэтому пусть $I_0\neq\phi$. Нулевой вектор согласно определениям соответствующих множеств принадлежит обоим конусам.

Пусть $z\in T(x^0)$, $z\neq 0$. Тогда существуют вектор $z^*\in T(x^0)$, $\|z^*\|=1$, и число $\lambda>0$ такие, что $z=\lambda z^*$, а также последовательность $\{x^k\}\subset G$, для которой $x^k\overset{z^*}{\to}x^0$. В силу того, что $g_i(x^k)\leqslant 0\ \ \forall i\in I$ и $g_i(x^0)=0$ $\forall i\in I_0$ имеем

$$\lim_{k\to\infty}\frac{g_i(x^k)-g_i(x^0)}{\|x^k-x^0\|}=\nabla g_i(x^0)'z^*\leqslant 0\quad\forall i\in I_0.$$

Следовательно, $z\in K(x^0)$.

Теорема 4.7. *Справедливо включение*

$$K^0(x^0)=\{z\in\mathbf{R}^n:\ \nabla g_{I_0}(x^0)'z<0\}\subseteq T(x^0).$$

Доказательство. В случаях $I_0=\phi$ и $K^0(x^0)=\phi$ утверждение теоремы очевидно, так что мы можем считать $I_0\neq\phi$ и $K^0(x^0)\neq\phi$. Пусть $z^*\in K^0(x^0)$. Без ограничения общности мы можем также считать, что $\|z^*\|=1$, так как для произвольного $z\in K^0(x^0)$ существуют $z^*\in K^0(x^0)$, $\|z^*\|=1$, и $\lambda>0$ такие, что $z=\lambda z^*$. Из разложения Тейлора

$$g_i(x^0+tz^*)-g_i(x^0)=t\nabla g_i(x^0)'z^*+o_i(t),\quad i\in I, \tag{4.9}$$

где $o_i(t)/t\to 0$ при $t\to 0$, вытекает, что для каждого $i\in I$ найдется число $t_i^0>0$ такое, что при $t\leqslant t_i^0$ выражение $t\nabla g_i(x^0)'z^*$ или равно нулю, или имеет тот же знак, что и правая часть в (4.9).

Для $i\in I_0$ в силу соотношений $\nabla g_i(x^0)'z^*<0$ и $g_i(x^0)=0$ следует

$$g_i(x^0+tz^*)<0\quad\forall t\leqslant t^1=\min_{i\in I_0}t_i^0. \tag{4.10}$$

Для $i\in I\setminus I_0$ справедливо $g_i(x^0)<0$. С учетом непрерывности g_i найдется число $t^2>0$ такое, что

$$g_i(x^0+tz^*)<0\quad\forall t\leqslant t^2. \tag{4.11}$$

Из (4.10) и (4.11) следует $g_i(x^0+tz^*)<0$ для всех $i\in I$ и $t\leqslant\min\{t^1,t^2\}$. Пусть $x^k=x^0+z^*/k$ для $k>1/\min\{t^1,t^2\}$. Тогда последовательность $\{x^k\}$ сходится к x^0 в направлении z^* и $x^k\in G$. Следовательно, $z^*\in T(x^0)$.

Как показывает следующий пример, соотношение $K(x^0) = T(x^0)$, вообще говоря, не выполняется.

Пример 4.2. Пусть $G = \{x \in \mathbf{R}^2 : -x_1^3 + x_2 \leqslant 0, -x_2 \leqslant 0\}$ и $x^0 = (0, 0)'$. Получаем (рис. 4.2)

$$K(x^0) = \{z \in \mathbf{R}^2 : z_2 = 0\},$$

$$T(x^0) = \{z \in \mathbf{R}^2 : z_1 \geqslant 0, z_2 = 0\}.$$

Таким образом, справедливо $T(x^0) \subset K(x^0)$.

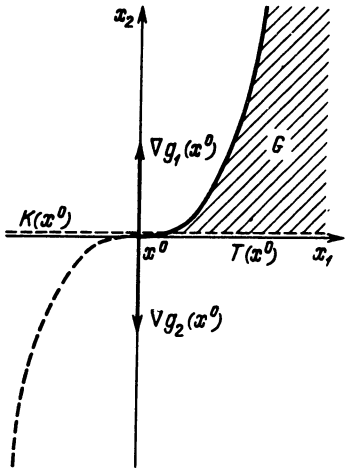


Рис. 4.2.

Этот пример показывает, в частности, что $K(x^0)$ не обязательно является "линейной аппроксимацией" множества G в точке $x^0 \in G$.

Линеаризирующий конус $K(x^0)$, в отличие от касательного конуса $T(x^0)$, зависит от аналитического представления G, что иллюстрирует следующий пример.

Пример 4.3. Множество G из примера 4.2 может быть представлено также в виде

$$G = \{x \in \mathbf{R}^2 : -x_1^3 + x_2 \leqslant 0, -x_2 \leqslant 0, -x_1 \leqslant 0\}.$$

В этом случае для $x^0 = (0, 0)'$ получается линеаризирующий конус

$$K(x^0) = \{z \in \mathbf{R}^n : z_1 \geqslant 0, z_2 = 0\},$$

и таким образом, $T(x^0) = K(x^0)$.

Пусть выполнены условия теоремы 4.3 и $I_0 \neq \phi$. Тогда в силу включения $K^0(x^0) \subseteq T(x^0)$ справедлива импликация

$$\nabla g_{I_0}(x^0)'z < 0 \Rightarrow \nabla f(x^0)'z \geqslant 0. \tag{4.12}$$

Эта импликация эквивалентна тому, что система неравенств

$$\nabla f(x^0)'z < 0, \quad \nabla g_{I_0}(x^0)'z < 0 \tag{4.13}$$

не имеет решения. По теореме 2.107 (теореме Жордана об альтернативе) система (4.13) тогда и только тогда неразрешима, когда система

$$(\nabla f(x^0), \nabla g_{I_0}(x^0))\begin{pmatrix} \eta \\ y_{I_0} \end{pmatrix} = 0, \quad \begin{pmatrix} \eta \\ y_{I_0} \end{pmatrix} \geqslant 0 \tag{4.14}$$

имеет нетривиальное решение. Положим далее

$$y_i = 0 \quad (i \in I \setminus I_0), \qquad y' = (y_1, \ldots, y_m),$$

$$\nabla g(x^0) = \begin{pmatrix} \partial g_1(x^0)/\partial x_1 & \ldots & \partial g_m(x^0)/\partial x_1 \\ \ldots\ldots\ldots\ldots\ldots & & \\ \partial g_1(x^0)/\partial x_n & \ldots & \partial g_m(x^0)/\partial x_n \end{pmatrix} = (\nabla g_1(x^0), \ldots, \nabla g_m(x^0)).$$

Тогда из существования нетривиального решения (4.14) вытекает разрешимость системы

$$\eta \nabla f(x^0) + \nabla g(x^0) y = 0, \quad g(x^0) \leqslant 0,$$
$$y' g(x^0) = 0, \quad (\eta, y)' \geqslant 0, \quad (\eta, y)' \neq 0.$$

Если $I_0 = \phi$ и функция f достигает в точке $x^0 \in G$ локального минимума на множестве G, то, так как $x^0 \in \operatorname{int} G$, эта система имеет решение, у которого $\eta > 0, y = 0$.

Из приведенных выше рассуждений вытекает критерий оптимальности, который был сформулирован в 1948 г. Джоном [42]. Рассмотрим задачу

P_2: найти $(x^0, y^0)' \in \mathbf{R}^n \times \mathbf{R}_+^m$ и $\eta^0 \in \mathbf{R}_+$, $\quad (\eta^0, y^0)' \neq 0$,

являющиеся решением следующей системы:

$$\eta \nabla f(x) + \nabla g(x) y = 0, \quad g(x) \leqslant 0,$$
$$y' g(x) = 0, \quad (\eta, y)' \geqslant 0. \tag{4.15}$$

Полагая

$$L_\eta(x, y) = \eta f(x) + y' g(x), \quad (x, y)' \in \mathbf{R}^n \times \mathbf{R}_+^m, \tag{4.16}$$

получим эквивалентную запись для системы (4.15):

$$\nabla_x L_\eta(x, y) = 0, \quad y' \nabla_y L_\eta(x, y) = 0,$$
$$\nabla_y L_\eta(x, y) \leqslant 0, \quad (\eta, y)' \geqslant 0. \tag{4.17}$$

Следующее утверждение характеризует связь между задачами P_1 и P_2.

Теорема 4.8 (*теорема Джона*). *Пусть* $f: \mathbf{R}^n \to \mathbf{R}$, $g: \mathbf{R}^n \to \mathbf{R}^m$ – *дифференцируемые функции и* $x^0 \in G$ – *точка локального минимума* f *на множестве* G. *Существуют* $y^0 \in \mathbf{R}_+^m$ *и* $\eta^0 \in \mathbf{R}_+$ *такие, что* $(x^0, y^0)'$ *и* η^0 *являются решением задачи* P_2.

Очевидно, что теорема 4.8 неприменима в качестве критерия оптимальности для случая $\eta^0 = 0$. Легко можно построить задачу P_1, для которой система

$$\nabla g_{I_0}(x^0)' z < 0 \tag{4.18}$$

неразрешима. Это происходит, например, в случае $K(x^0) = \phi$. По теореме 2.107 из неразрешимости (4.18) следует, что система

$$\nabla g_{I_0}(x^0) y_{I_0} = 0, \quad y_{I_0} \geqslant 0, \tag{4.19}$$

имеет нетривиальное решение $y_{I_0}^0$. Если положить $y_i^0 = 0$ для $i \in I \setminus I_0$, то вектор $(0, y^0)'$, где $y^0 = (y_1^0, \ldots, y_m)'$, будет нетривиальным решением системы (4.15), не зависящим от значения $f(x^0)$ и от поведения функции f в окрестности точки x^0.

Дополнительное условие, которое при предположениях теоремы 4.8 обеспечивает для любого решения задачи P_2 выполнение соотношения $\eta^0 > 0$, называется *условием регулярности*.

Из доказательства теоремы 4.8 напрашивается в качестве условия регулярности требование $K^0(x^0) \neq \phi$. В этом случае система (4.18) имеет ре-

шение, следовательно, у (4.19) отсутствуют нетривиальные решения. Поэтому в (4.15) должно быть $\eta^0 \neq 0$. К условию $K^0(x^0) \neq \phi$ мы вернемся в п.4.2.3.

Менее жесткое ограничение на допустимый класс функций для задачи P_2 накладывает следующее условие регулярности:

R_1: функция g: $\mathbf{R}^n \to \mathbf{R}^m$ дифференцируема и имеет место равенство $K(x^0) = T(x^0)$.

Для задачи с одними лишь аффинными ограничениями условие регулярности R_1 всегда выполнено.

Учитывая сделанные ранее замечания, следует еще раз подчеркнуть, что R_1 является условием на аналитическое представление множества G. Как показывают примеры 4.2 и 4.3, для одного и того же множества G условие R_1 в одном случае выполняется, а в другом нет.

Следующая основная теорема теории необходимых локальных критериев оптимальности была опубликована в 1951 г. Куном и Таккером [50].

По аналогии с P_2 рассмотрим задачу

P_3 : найти точку $(x^0, y^0)' \in \mathbf{R}^n \times \mathbf{R}^m_+$, являющуюся решением следующей системы:

$$\nabla f(x) + \nabla g(x) y = 0, \quad y' g(x) = 0, \tag{4.20}$$

$$g(x) \leqslant 0, \quad y \geqslant 0. \tag{4.21}$$

Соотношения (4.20), (4.21) называются (*локальными*) *условиями Куна – Таккера.*

Применяя *функцию Лагранжа*

$$L(x, y) = f(x) + y' g(x), \quad (x, y)' \in \mathbf{R}^n \times \mathbf{R}^m_+, \tag{4.22}$$

получим для системы (4.20), (4.21) эквивалентную формулировку:

$$\begin{aligned} &\nabla_x L(x, y) = 0, \quad y' \nabla_y L(x, y) = 0, \\ &\nabla_y L(x, y) \leqslant 0, \quad y \geqslant 0. \end{aligned} \tag{4.23}$$

Функция Лагранжа L получается из определенной в (4.16) функции L_η при $\eta = 1$. В этом смысле L_η является обобщением функции Лагранжа. Связь между задачами P_1 и P_3 характеризуется следующим утверждением.

Т е о р е м а 4.9 (*теорема Куна – Таккера*). *Пусть* f: $\mathbf{R}^n \to \mathbf{R}$, g: $\mathbf{R}^n \to \mathbf{R}^m$ *– дифференцируемые функции,* $x^0 \in G$ *– точка локального минимума* f *на множестве* G *и в* x^0 *выполнено условие регулярности* R_1. *Существует вектор* $y^0 \in \mathbf{R}^m_+$ *такой, что точка* $(x^0, y^0)'$ *будет решением задачи* P_3.

Д о к а з а т е л ь с т в о. Так как условие R_1 выполнено, то из теоремы 4.3 для $I_0 \neq \phi$ тотчас следует

$$\nabla g_{I_0}(x^0)' z \leqslant 0 \Rightarrow \nabla f(x^0)' z \geqslant 0. \tag{4.24}$$

Применяя теорему 2.43 (теорему Фаркаша), получаем, что соотношение (4.24) имеет место тогда и только тогда, когда разрешима система

$$\nabla f(x^0) + \nabla g_{I_0}(x^0) y_{I_0} = 0, \quad y_{I_0} \geqslant 0. \tag{4.25}$$

Полагая дополнительно $y_i = 0 \ \forall i \in I \setminus I_0$, видим, что теорема справедлива для $I_0 \neq \phi$. В случае $I_0 = \phi$ очевидным решением задачи P_3 является $y^0 = 0$.

Замечания. 1. Если $x^0 \in \operatorname{int} G$, то множество I_0 пусто (в силу R_1), и утверждение теоремы 4.9 совпадает с теоремой 1.1.

2. При применении теоремы Куна – Таккера условие регулярности R_1 часто заменяется одним из приведенных ниже (в п.4.2.3) условий $R_2 - R_7$. Каждое из этих условий является достаточным для выполнения R_1.

Утверждение теоремы 4.9 допускает простую геометрическую интерпретацию. Соотношение (4.25) при $I_0 \neq \phi$ означает, что вектор $-\nabla f(x^0)$ является элементом многогранного конуса

$$C = \{u \in \mathbf{R}^n: \ u = \nabla g_{I_0}(x^0) y_{I_0}, \quad y_{I_0} \geqslant 0\},$$

образованного градиентами активных в точке x^0 ограничений (рис. 4.3).

По теореме Фаркаша это эквивалентно следующему утверждению: отрицательный полярный конус

$$-C^+ = K(x^0) = \{z \in \mathbf{R}^n: \ \nabla g_{I_0}(x^0)' z \leqslant 0\}$$

лежит в полупространстве $E_+(\nabla f(x^0), 0) = \{z \in \mathbf{R}^n: \nabla f(x^0)' z \geqslant 0\}$.

На рис. 4.3 в точке x^0 имеются два активных ограничения, а на рис. 4.4 показан случай лишь одного активного в x^0 ограничения. Во втором случае множество C есть луч, а $K(x^0)$ – полупространство.

Справедливость соотношения (4.24) может быть проверена с помощью некоторой задачи линейного программирования. Импликация (4.24) выполняется тогда и только тогда, когда оптимальное значение задачи

$$\min\{\nabla f(x^0)' z: \ \nabla g_{I_0}(x^0)' z \leqslant 0\} \tag{4.26}$$

равно нулю.

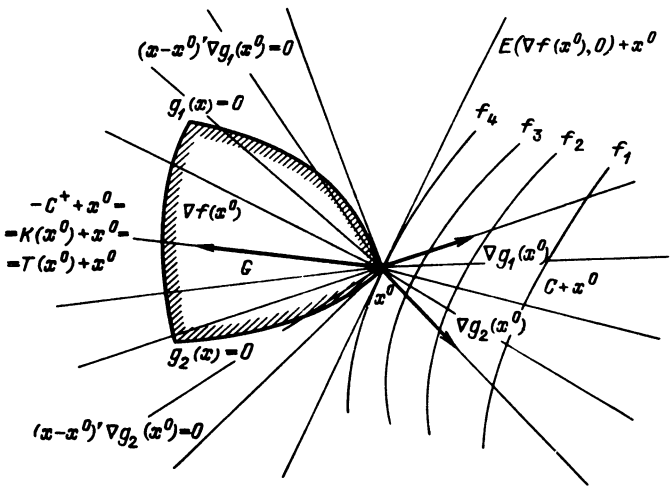


Рис. 4.3.

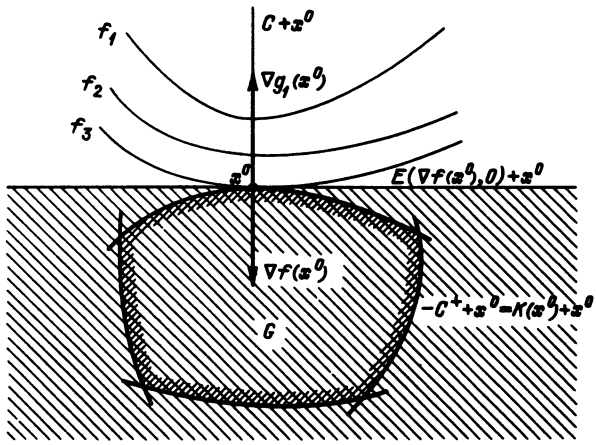


Рис. 4.4.

Пример 4.4. (*Клейнмихель* [46]). Рассмотрим задачу

$$\min\{x_1 + x_2 : x = (x_1, x_2)' \in G \subseteq \mathbb{R}^2\},$$

$$G = \{x = (x_1, x_2)' : g_1(x) = -1 - x_1^3 + x_2 \leqslant 0, \quad g_2(x) = -1 + x_1 \leqslant 0,$$

$$g_3(x) = -x_2 \leqslant 0\}, \tag{4.27}$$

и исследуем, являются ли ее решениями точки $x^0 = (-1, 0)'$ и $\bar{x}^0 = (0, 1)'$.

Для x^0 соответствующая (4.26) задача

$$\min\left\{z_1 + z_2 : \begin{pmatrix} -3 & 1 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \leqslant 0\right\} \tag{4.28}$$

имеет оптимальное значение 0.

Для $\bar{x}^0$ задаче (4.26) соответствует задача

$$\min\left\{z_1 + z_2 : (0, 1)\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \leqslant 0\right\}, \tag{4.29}$$

целевая функция которой не ограничена снизу на допустимом множестве. Поэтому x^0 может быть решением задачи (4.27), а $\bar{x}^0$ – нет.

В предыдущих рассуждениях о критериях оптимальности предполагалось существование лишь первых производных рассматриваемых функций. Поэтому об этих критериях говорят как об условиях первого порядка. Для дважды дифференцируемых функций могут быть сформулированы так называемые условия второго порядка (см. также [27]).

Пусть $x^0 \in G$ и $(x^0, y^0)'$ – решение задачи P_3. Это решение определяет множество индексов $I_1 = \{i \in I_0 : y_i^0 > 0\}$.

Сформулируем два условия:

B_1: касательный конус $T_1(x^0)$ к множеству

$$G_1 = G \cap \{x : g_i(x) = 0 \quad \forall i \in I_1\}$$

совпадает с линеаризирующим конусом

$$K_1(x^0) = \{ z: \nabla g_i(x^0)'z \leqslant 0 \quad \forall i \in I_0 \setminus I_1, \quad \nabla g_i(x^0)'z = 0 \quad \forall i \in I_1 \};$$

B_2: векторы $\nabla g_i(x^0)$, $i \in I_1$, линейно независимы и существует точка $z \in \mathbf{R}^n$ такая, что

$$\nabla g_i(x^0)'z < 0 \quad \forall i \in I_0 \setminus I_1, \quad \nabla g_i(x^0)'z = 0 \quad \forall i \in I_1.$$

Имеет место следующий необходимый критерий оптимальности для задачи P_1.

Теорема 4.10. *Пусть* $f: \mathbf{R}^n \to \mathbf{R}$, $g: \mathbf{R}^n \to \mathbf{R}^m$ *– дважды дифференцируемые в* $x^0 \in G$ *функции,* x^0 *– точка локального минимума* f *на множестве* G *и* $(x^0, y^0)'$ *– решение задачи* P_3. *Если в* x^0 *выполнены условие регулярности* R_1 *и условие* B_1, *то*

$$z' \nabla_x^2 L(x^0, y^0) z \geqslant 0 \quad \forall z \in K_1(x^0).$$

Доказательство. Для $x \in G_1$, очевидно, имеем $L(x, y^0) = f(x)$.

Применяя теорему 4.4 к задаче P_1 с допустимой областью G_1 вместо G, получим

$$z' \nabla_x^2 L(x^0, y^0) z \geqslant 0 \quad \forall z \in T_1(x^0).$$

С учетом условия B_1 отсюда следует утверждение теоремы.

Пример 4.5. Для задачи

$$\min\{ f(x) = -x_1^2 \omega - x_2 : \; x = (x_1, x_2)' \in G \subseteq \mathbf{R}^2 \},$$

$$G = \{ x: \; g_1(x) = -x_1^2 - (x_2 - 1)^2 + 1 \leqslant 0,$$

$$g_2(x) = (x_1 + 1)^2 + x_2^2 - 1 \leqslant 0, \quad g_3(x) = -x_1 - 1 \leqslant 0 \},$$

где $\omega \in \mathbf{R}$ – параметр, исследуем, удовлетворяет ли $x^0 = (0, 0)'$ необходимым условиям теоремы 4.9. У нас $I_0 = \{1, 2\}$, $\nabla f(x^0) = (0, -1)'$, $\nabla g_1(x^0) = (0, 2)'$, $\nabla g_2(x^0) = (2, 0)'$, $\nabla g_3(x^0) = (-1, 0)'$. Кроме того, выполнено условие регулярности R_4 (см. п.4.2.3).

Видно, что $y^0 = (1/2, 0, 0)'$ и $x^0 = (0, 0)'$ решают P_3:

$$\begin{pmatrix} 0 \\ -1 \end{pmatrix} + \begin{pmatrix} 0 & 2 & -1 \\ 2 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_1^0 \\ y_2^0 \\ y_3^0 \end{pmatrix} = 0.$$

Данное необходимое условие выполняется независимо от выбора параметра ω, хотя очевидно, что для $\omega = 0$ точка $(0, 0)'$ не является решением данной задачи. Приведенный в теореме 4.10 критерий оптимальности не выполняется для $\omega < 1/2$.

Для применения теоремы 4.10 нам потребуются легко проверяемые условия, достаточные для выполнения B_1. Одним из таких условий является B_2.

Теорема 4.11. *Справедлива импликация* $B_2 \Rightarrow B_1$.

Доказательство. Пусть вначале $I_1 \neq \phi$. Так как векторы $\nabla g_i(x^0)$, $i \in I_1$, линейно независимы, то существует матрица A порядка $n \times |I_1|$, для которой

$$|\nabla g_{I_1}(x^0)'A| \neq 0. \tag{4.30}$$

Рассмотрим вектор $z \in \mathbf{R}^n$, $z \neq 0$, удовлетворяющий условию B_2, а также систему уравнений относительно y_{I_1} и t вида

$$g_i(x^0 + Ay_{I_1} + tz) = 0, \quad i \in I_1. \tag{4.31}$$

Система (4.31) имеет решение $(y_{I_1}, t^0)' = (0, 0)'$. В силу (4.30) по теореме о разрешимости (теорема 1.2) в некоторой окрестности U_0 точки $t^0 = 0$ существуют функции $y_i(t), i \in I_1$, такие, что

$$y_i(0) = 0, \quad i \in I_1,$$

$$g_i(x^0 + Ay_{I_1}(t) + tz) = 0, \quad i \in I_1, \quad t \in U_0. \tag{4.32}$$

Дифференцируя (4.32) по t в точке $t = 0$, получим

$$\nabla g_i(x^0)'(A\dot{y}_{I_1}(0) + z) = 0, \quad i \in I_1.$$

Из $\nabla g_i(x^0)'z = 0 \ \forall i \in I_1$ следует, что

$$\nabla g_i(x^0)'A\dot{y}_{I_1}(0) = 0, \quad i \in I_1.$$

Учитывая (4.30), имеем $\dot{y}_{I_1}(0) = 0$ и тем самым

$$\lim_{t \to 0} \frac{y_{I_1}(t)}{t} = 0. \tag{4.33}$$

Покажем теперь, что найдется число $t_1 > 0$ такое, что кривая

$$x(t) = x^0 + Ay_{I_1}(t) + tz, \quad t \in [0, t_1],$$

полностью лежит в множестве G_1. Для этого мы разложим функции g_i, $i \in I_0 \setminus I_1$, в точке x^0 по $Ay_{I_1}(t) + tz$ и учтем равенства $g_i(x^0) = 0$. Получим

$$g_i(x^0 + Ay_{I_1}(t) + tz) = \nabla g_i(x^0)'(Ay_{I_1}(t) + tz) + o(\| Ay_{I_1}(t) + tz \|) =$$

$$= t\nabla g_i(x^0)'z + o(\| Ay_{I_1}(t) + tz \|) + \nabla g_i(x^0)'Ay_{I_1}(t).$$

В силу (4.33) имеем $\nabla g_i(x^0)'Ay_{I_1}(t) = o(t)$ и

$$g_i(x^0 + Ay_{I_1}(t) + tz) = t\nabla g_i(x^0)'z + o(t).$$

Поскольку вектор z удовлетворяет условию B_2, то $\nabla g_i(x^0)'z < 0$ для $i \in I_0 \setminus I_1$. Поэтому существует такое число $t_0 > 0$, что для $t \in (0, t_0)$ справедливо

$$g_i(x^0 + Ay_{I_1}(t) + tz) < 0 \quad \forall i \in I_0 \setminus I_1.$$

Отсюда и из (4.32) вытекает существование числа $t_1 > 0$, при котором кривая $x(t), t \in [0, t_1]$, полностью содержится в G_1.

Теперь мы можем задать последовательность $x^k \in G_1$ со свойством $x^k \xrightarrow{z \| z \|^{-1}} x^0$, полагая $x^k = x^0 + Ay_{I_1}(1/k) + z/k$. В самом деле, при

$k > 1/t_1$ имеем $x^k \in G_1$ и в силу (4.33)

$$\lim_{k\to\infty} \frac{x^k - x^0}{\| x^k - x^0 \|} = \lim_{k\to\infty} \frac{Ay_{I_1}(1/k) + z/k}{\| Ay_{I_1}(1/k) + z/k \|} =$$

$$= \lim_{k\to\infty} \frac{Ay_{I_1}(1/k)k + z}{\| Ay_{I_1}(1/k)k + z \|} = \frac{z}{\| z \|},$$

откуда следует утверждение теоремы для $I_1 \neq \phi$. Доказательство для случая $I_1 = \phi$ содержится в доказательстве теоремы 4.17 (для этого в последнем следует положить $I_0^N = I_0$ и $I_0^L = \phi$).

Пример 4.6. Рассмотрим задачу

$$\min\{ f(x) = \omega x_1^2 : \ x = (x_1, x_2)' \in G \subseteq \mathbf{R}^2 \},$$

$$G = \{ x \in \mathbf{R}^2 : \ g_1(x) = (x_1 + 1)^2 + x_2^2 - 1 \leqslant 0,$$

$$g_2(x) = x_2 \leqslant 0, \quad g_3(x) = -x_2 \leqslant 0 \},$$

где $\omega \in \mathbf{R}$ – параметр. Для $x^0 = (0, 0)'$ имеем $I_0 = \{1, 2, 3\}$, $\nabla f(x^0) = (0, 0)'$, $\nabla g_1(x^0) = (2, 0)'$, $\nabla g_2(x^0) = (0, 1)'$, $\nabla g_3(x^0) = (0, -1)'$.

Нетрудно видеть, что $y^0 = (0, 1, 1)'$ и $x^0 = (0, 0)'$ решают задачу P_3 с $I_1 = \{2, 3\}$. В силу того, что $K(x^0) = T(x^0) = \{ z \in \mathbf{R}^2 : \ z_1 \leqslant 0, z_2 = 0 \}$, выполнено и условие регулярности R_1.

Допустим теперь, что $x^0 = (0, 0)'$ – точка локального минимума функции f на множестве G. В силу соотношений $G_1 = G$ и $K_1(x^0) = K(x^0)$ условие B_1 выполнено. Тогда по теореме 4.10

$$z' \begin{pmatrix} 2\omega & 0 \\ 0 & 0 \end{pmatrix} z \geqslant 0 \qquad \forall z \in K_1(x^0).$$

Отсюда следует $2\omega z_1^2 \geqslant 0 \ \ \forall z_1 \leqslant 0$ и, таким образом, $\omega \geqslant 0$.

Теорема 4.12. *Пусть* $f: \mathbf{R}^n \to \mathbf{R}$, $g: \mathbf{R}^n \to \mathbf{R}^m$ – *дважды дифференцируемые в* $x^0 \in G$ *функции. Предположим, что существует точка* $(x^0, y^0)' \in \mathbf{R}^n \times \mathbf{R}_+^m$ *такая, что* $D(x^0) = T(x^0) \cap \{ z : \nabla g_i(x^0)' z = 0 \ \forall i \in I_1 \} \neq \phi$ *и выполняются условия*

$$\nabla_x L(x^0, y^0) = 0, \quad g(x^0) \leqslant 0, \quad y^{0\prime} g(x^0) = 0, \quad y^0 \geqslant 0, \tag{4.34а}$$

$$z' \nabla_x^2 L(x^0, y^0) z > 0 \quad \forall z \in D(x^0), \ z \neq 0, \tag{4.34б}$$

где $I_1 = \{ i \in I_0 : \ y_i^0 > 0 \}$. *Тогда* x^0 – *точка собственного локального минимума* f *на множестве* G.

Доказательство. Предположим противное. Пусть функция f не достигает в точке x^0 собственного локального минимума на множестве G. Тогда существует последовательность $\{x^k\} \subset G$ такая, что $x^k \overset{z}{\to} x^0$ и $f(x^k) - f(x^0) < 0$ для $k \in \mathbf{N}$. Отсюда следует, что

$$\lim_{k\to\infty} \frac{f(x^k) - f(x^0)}{\| x^k - x^0 \|} = \nabla f(x^0)' z \leqslant 0,$$

и, поскольку $g_i(x^k) \leqslant 0$, $g_i(x^0) = 0$ $\forall i \in I_0$, то

$$\lim_{k \to \infty} \frac{g_i(x^k) - g_i(x^0)}{\| x^k - x^0 \|} = \nabla g_i(x^0)' z \leqslant 0, \quad i \in I_0.$$

Тем самым мы получаем неравенство

$$L(x^k, y^0) - L(x^0, y^0) = f(x^k) - f(x^0) + \sum_{i \in I_0} y_i^0 [g_i(x^k) - g_i(x^0)] \leqslant 0.$$

Из равенства $\nabla_x L(x^0, y^0) = 0$ и теоремы 4.1 следует, что

$$\lim_{k \to \infty} \frac{L(x^k, y^0) - L(x^0, y^0)}{\| x^k - x^0 \|^2} = \frac{1}{2} z' \nabla_x^2 L(x^0, y^0) z \leqslant 0.$$

По предположению $z \in T(x^0)$. Если $I_1 = \phi$, то сразу же получаем противоречие с (4.34б). Если $I_1 \neq \phi$, то допустим, что $\nabla g_{i^*}(x^0)' z < 0$ для некоторого $i^* \in I_1$. Умножая скалярно $\nabla_x L(x^0, y^0)$ на z и учитывая, что $\nabla f(x^0)' z \leqslant 0$ и $y_i^0 > 0$ $\forall i \in I_1$, имеем

$$0 = \nabla_x L(x^0, y^0)' z = \nabla f(x^0)' z + \sum_{i \in I_1} y_i^0 \nabla g_i(x^0)' z < 0.$$

Получили противоречие. Следовательно, $\nabla g_i(x^0)' z = 0$ $\forall i \in I_1$.

Таким образом, предположение о том, что функция f не достигает в x^0 локального минимума на G, противоречит (4.34б). Поэтому каждая точка $(x^0, y^0)' \in \mathbf{R}^n \times \mathbf{R}_+^m$, удовлетворяющая условиям (4.34), определяет точку x^0, в которой функция f достигает локального минимума на множестве G.

В силу того, что $T(x^0) \subseteq K(x^0)$, теорему 4.12 можно уточнить, а именно, потребовать, чтобы выполнялись неравенства

$$z' \nabla_x^2 L(x^0, y^0) z > 0 \qquad \forall z \in K_1(x^0), \quad z \neq 0,$$

где

$$K_1(x^0) = \{ z: \nabla g_i(x^0)' z \leqslant 0 \quad \forall i \in I_0 \setminus I_1, \ \nabla g_i(x^0)' z = 0 \quad \forall i \in I_1 \}.$$

Пример 4.7. Обратимся еще раз к задаче из примера 4.5. Для $\omega > 1/2$ и $z_1 > 0$ теорема 4.12, очевидно, справедлива, т.е. при $\omega > 1/2$ функция f достигает в $x^0 = (0, 0)'$ локального минимума на множестве G. При $\omega = 1/2$ имеем $z' \nabla_x^2 L(x^0, y^0) z = 0$, так что аналогичное утверждение невозможно.

4.2.3. Условия регулярности для задачи P_1. В 4.2.2 было введено условие регулярности R_1. Теперь мы приведем другие условия регулярности и исследуем существующие между ними связи. При этом оказывается, что R_1 является слабейшим условием в том смысле, что все другие вводимые здесь условия влекут за собой R_1. Преимуществом нижеследующих условий по сравнению с R_1 является более легкая проверка их выполнения. Поэтому при применении необходимого критерия оптимальности Куна – Таккера (теорема 4.9) требование выполнения условия R_1 заменяется требованием выполнения одного из следующих ниже условий.

В дальнейшем I^L – множество всех $i \in I$, для которых g_i – аффинная функция, I^N – множество всех $i \in I$, для которых g_i не является аффинной функцией; $I = I^L \cup I^N$, $I^L \cap I^N = \phi$. Множество индексов I_0 активных ограничений представляется соответственно как $I_0 = I_0^L \cup I_0^N$.

Сформулируем ряд условий регулярности (предполагается, что $x^0 \in G$):

R_1: функция g: $\mathbf{R}^n \to \mathbf{R}^m$ дифференцируема и выполняется равенство

$$T(x^0) = K(x^0);$$

R_2: функция g: $\mathbf{R}^n \to \mathbf{R}^m$ дифференцируема и система

$$\sum_{i \in I_0^N} y_i \nabla g_i(x^0) + \sum_{i \in I_0^L} y_i \nabla g_i(x^0) = 0;$$
$$y_i \geqslant 0 \quad \forall i \in I_0^N \cup I_0^L; \qquad \exists i \in I_0^N: \ y_i > 0, \tag{4.35а}$$

не имеет решений;

R_3: функция g: $\mathbf{R}^n \to \mathbf{R}^m$ дифференцируема и существует вектор $z \in \mathbf{R}^n$ такой, что

$$\nabla g_i(x^0)'z < 0 \quad \forall i \in I_0^N;$$
$$\nabla g_i(x^0)'z < 0 \quad \forall i \in I_0^L; \tag{4.35б}$$

R_4: функция g: $\mathbf{R}^n \to \mathbf{R}^m$ дифференцируема и система

$$\sum_{i \in I_0^N} y_i \nabla g_i(x^0) + \sum_{i \in I_0^L} y_i \nabla g_i(x^0) = 0;$$
$$\exists i \in I_0^N: \ y_i \neq 0,$$

не имеет решений;

R_5: функции g_i: $\mathbf{R}^n \to \mathbf{R}$, $i \in I^N$, выпуклы и существует вектор $x^1 \in G$ такой, что

$$g_i(x^1) < 0 \quad \forall i \in I^N; \qquad g_i(x^1) \leqslant 0 \quad \forall i \in I^L;$$

R_6: функции g_i: $\mathbf{R}^n \to \mathbf{R}$, $i \in I^N$, выпуклы и для любого $i \in I^N$ найдется вектор $x^i \in G$ со свойством $g_i(x^i) < 0$;

R_7: функции g_i: $\mathbf{R}^n \to \mathbf{R}$, $i \in I^N$, квазивыпуклы и существует вектор $x^1 \in G$, для которого

$$g_i(x^1) < 0 \quad \forall i \in I^N; \quad g_i(x^1) \leqslant 0 \quad \forall i \in I^L.$$

Теорема 4.13. *Условия R_2 и R_3 эквивалентны.*

Доказательство. По теореме Моцкина об альтернативе (теорема 2.106) система (4.35а) не имеет решений тогда и только тогда, когда система (4.35б) разрешима в точке z.

Теорема 4.14. *Условие R_4 влечет условие R_2.*

Доказательство. Если выполняется R_4, то неразрешима и система (4.35а), т.е. имеет место условие R_2.

Теорема 4.15. *Пусть f: $\mathbf{R}^n \to \mathbf{R}$, g: $\mathbf{R}^n \to \mathbf{R}^m$ – дифференцируемые, а g_i, $i \in I^N$, – выпуклые функции. Тогда условия R_3 и R_5 эквивалентны.*

Доказательство. 1. Пусть выполнено условие R_5. Поскольку функции g_i, $i \in I^N$, выпуклые, то для $x^0 \in G$ в силу равенств $g_i(x^0) = 0$

$\forall i \in I_0$ имеем

$$(x^1 - x^0)' \nabla g_i(x^0) \leqslant g_i(x^1) - g_i(x^0) < 0 \quad \forall i \in I_0^N, \tag{4.36}$$

$$(x^1 - x^0)' \nabla g_i(x^0) = g_i(x^1) - g_i(x^0) \leqslant 0 \quad \forall i \in I_0^L. \tag{4.37}$$

Очевидно, что условие R_3 выполняется при $z = x^1 - x^0$.

2. Пусть выполнено условие R_3. Тогда для любых x^0, $x^1 \in \mathbf{R}^n$ справедливы неравенства

$$(x^1 - x^0)' \nabla g_i(x^0) \leqslant g_i(x^1) - g_i(x^0) \quad \forall i \in I^N,$$

$$(x^1 - x^0)' \nabla g_i(x^0) = g_i(x^1) - g_i(x^0) \quad \forall i \in I^L.$$

В частности, полагая $x^0 \in G$, $x^1 = x^0 + \lambda z$ и учитывая R_3, непрерывность g_i, $i \in I$, и соотношения $g_i(x^0) = 0$ $\forall i \in I_0$, $g_i(x^0) < 0$ $\forall i \in I \setminus I_0$, для достаточно малого $\lambda > 0$ получим $g_i(x^0 + \lambda z) < 0$ $\forall i \in I_0^N$, $g_i(x^0 + \lambda z) \leqslant 0$ $\forall i \in I_0^L$, $g_i(x^0 + \lambda z) < 0$ $\forall i \in I \setminus I_0$. Тем самым условие R_5 выполнено.

Для выпуклых функций импликация $R_7 \Rightarrow R_5$ тривиальна.

Условия R_5 – R_7 имеют значение и для глобальной теории множителей Лагранжа, в рамках которой требуется не дифференцируемость, а выпуклость функций f и g.

Теорема 4.16. *Условия R_6 и R_5 эквивалентны.*

Доказательство. 1. Соотношение $R_5 \Rightarrow R_6$ очевидно.

2. Пусть выполнено R_6. Тогда для

$$x^1 = \sum_{i \in I^N} \lambda_i x^i, \quad \lambda_i \in (0, 1), \quad \sum_{i \in I^N} \lambda_i = 1,$$

и произвольного фиксированного $i^* \in I^N$ имеем

$$g_{i^*}(x^1) = g_{i^*}\Big(\sum_{i \in I^N} \lambda_i x^i\Big) \leqslant \sum_{i \in I^N} \lambda_i g_{i^*}(x^i) \leqslant \lambda_{i^*} g_{i^*}(x^{i^*}) < 0.$$

Таким образом, $g_i(x^1) < 0$ $\forall i \in I^N$. Так как $x^1 \in G$, то $g_i(x^1) \leqslant 0$ $\forall i \in I^L$.

Теорема 4.17. *Условие R_3 влечет условие R_1.*

Доказательство. Пусть z^1 – произвольный элемент из конуса $K(x^0)$, вектор z, $z \neq 0$, удовлетворяет (4.35б). Тогда для всех достаточно малых $\lambda > 0$ имеем

$$\nabla g_i(x^0)'(z^1 + \lambda z) < 0 \quad \forall i \in I_0^N,$$

$$\nabla g_i(x^0)'(z^1 + \lambda z) \leqslant 0 \quad \forall i \in I_0^L, \quad z^1 + \lambda z \neq 0.$$

Как и в доказательстве теоремы 4.7, найдется число $t^0 > 0$ такое, что

$$x^0 + t \frac{z^1 + \lambda z}{\| z^1 + \lambda z \|} \in G \quad \forall t \in [0, t^0].$$

Положим $x^k = x^0 + \frac{1}{k} \frac{z^1 + \lambda z}{\| z^1 + \lambda z \|}$. Последовательность $\{x^k\}$ сходится к x^0 в направлении $\frac{z^1 + \lambda z}{\| z^1 + \lambda z \|}$; при этом $x^k \in G$ для $k > 1/t^0$.

Следовательно, $\dfrac{z^1 + \lambda z}{\| z^1 + \lambda z \|} \in T(x^0)$, а тем самым и $z^1 + \lambda z \in T(x^0)$. Поскольку $\lambda > 0$ можно выбрать сколь угодно малым, а множество $T(x^0)$ замкнуто, то $z^1 \in T(x^0)$, т.е. $K(x^0) \subseteq T(x^0)$. Отсюда в силу теоремы 4.16 следует справедливость R_1.

В заключение покажем на примере, что условие R_1 не является необходимым для того, чтобы задача P_3 имела решение.

Пример 4.8. Пусть $f(x) = x_2$, $x \in \boldsymbol{R}^2$, и, как в примере 4.2,

$$G = \{x \in R^2 : g_1(x) = -x_1^3 + x_2 \leqslant 0,\ g_2(x) = -x_2 \leqslant 0\}.$$

Очевидно, $x^0 = (0, 0)'$ – решение задачи P_1. Кроме того, $y^0 = (0,1)'$ вместе с x^0 является решением задачи P_3, хотя условие R_1 и не выполнено. Тем самым теорема 4.9 снова применима, если представить множество G так, как это сделано в примере 4.3.

Задача

$$\min \{x_1 + x_2 : -x_1^3 + x_2 \leqslant 0,\ -x_2 \leqslant 0\} \tag{4.38}$$

есть пример того, что P_2 может иметь решение (с $\eta^0 = 0$), в то время как P_3 неразрешима. Если представить допустимую область задачи (4.38) так же, как в примере 4.3, то P_2 будет иметь решение с $\eta^0 > 0$, а поэтому будет разрешима и задача P_3.

4.2.4. Локальная теория множителей Лагранжа для задачи P_1^*. При помощи полученных в п. 4.2.2, 4.2.3 результатов утверждения о задаче P_1 могут быть перенесены на случай задачи P_1^*. При этом мы ограничимся аналогом теоремы 4.5, исследованием условий регулярности и одним достаточным критерием оптимальности.

Функцией Лагранжа L^* для задачи P_1^* назовем функцию

$$L^*(x', y, v) = f(x) + y'g(x) + v'h(x), \quad (x, y, v)' \in \mathbf{R}^n \times \mathbf{R}_+^m \times \mathbf{R}^p. \tag{4.39}$$

По аналогии с P_3 рассмотрим задачу

P_3^*: найти точку $(x^0, y^0, v^0)' \in \mathbf{R}^n \times \mathbf{R}_+^m \times \mathbf{R}^p$, являющуюся решением следующей системы:

$$\begin{aligned} &\nabla f(x) + \nabla g(x)'y + \nabla h(x)'v = 0,\ g(x) \leqslant 0, \\ &y'g(x) = 0,\ y \geqslant 0, \\ &h(x) = 0. \end{aligned} \tag{4.40}$$

Система (4.40) может быть записана также в виде

$$\begin{aligned} &\nabla_x L^*(x, y, v) = 0,\ y'\nabla_y L^*(x, y, v) = 0, \\ &\nabla_y L^*(x, y, v) \leqslant 0,\ y \geqslant 0, \\ &\nabla_v L^*(x, y, v) = 0. \end{aligned} \tag{4.41}$$

Для формулировки необходимых критериев оптимальности нам снова потребуются некоторые условия регулярности. С этой целью для точки $x^0 \in G^*$ в соответствии с (4.5) определяются касательный конус $T^*(x^0)$ и линеаризирующий конус

$$K^*(x^0) = \{z \in \mathbf{R}^n : \nabla g_i(x^0)'z \leqslant 0,\ i \in I_0;\ \nabla h_j(x^0)'z = 0,\ j \in I^*\}.$$

Через I_0 обозначим как и прежде, множество индексов активных в точке x^0 ограничений. Сформулируем первое условие регулярности:

R_1^*: функции $g\colon \mathbf{R}^n \to \mathbf{R}^m$, $h\colon \mathbf{R}^n \to \mathbf{R}^p$, дифференцируемы и справедливо равенство $K^*(x^0) = T^*(x^0)$.

Применяя теоремы 2.43 и 4.3, получим следующее утверждение.

Теорема 4.18. *Пусть* $f\colon \mathbf{R}^n \to \mathbf{R}$, $g\colon \mathbf{R}^n \to \mathbf{R}^m$, $h\colon \mathbf{R}^n \to \mathbf{R}^p$ – *дифференцируемые функции,* $x^0 \in G^*$ – *точка локального минимума* f *на множестве* G^* *и в* x^0 *выполнено условие регулярности* R_1^*. *Тогда существует вектор* $(y^0, v^0)' \in \mathbf{R}_+^m \times \mathbf{R}^p$ *такой, что* $(x^0, y^0, v^0)'$ *является решением задачи* P_3^*.

Для практического использования условие R_1^* не особенно удачно. Поэтому, как и в п. 4.2.3, мы приведем ряд условий регулярности, каждое из которых влечет за собой R_1^*:

R_2^*: функции $g\colon \mathbf{R}^n \to \mathbf{R}^m$, $h\colon \mathbf{R}^n \to \mathbf{R}^p$ дифференцируемы и системы

$$\sum_{i \in I_0} y_i \nabla g_i(x^0) + \sum_{j \in I^*} y_j \nabla h_j(x^0) = 0,$$

$$y_i \geqslant 0 \quad \forall i \in I_0, \qquad \exists i \in I_0\colon\ y_i > 0,$$

$$\sum_{j \in I^*} y_j \nabla h_j(x^0) = 0, \quad y_j \neq 0,$$

не имеют решения;

R_3^*: функции $g\colon \mathbf{R}^n \to \mathbf{R}^m$, $h\colon \mathbf{R}^n \to \mathbf{R}^p$ дифференцируемы, существует вектор $z \in \mathbf{R}^n$ такой, что

$$\nabla g_i(x^0)'z < 0 \quad \forall i \in I_0, \qquad \nabla h_j(x^0)'z = 0 \quad \forall j \in I^*,$$

и векторы $\nabla h_j(x^0)$, $j \in I^*$, линейно независимы;

R_4^*: функции $g\colon \mathbf{R}^n \to \mathbf{R}^m$, $h\colon \mathbf{R}^n \to \mathbf{R}^p$ дифференцируемы и градиенты

$$\nabla g_i(x^0), \quad i \in I_0, \qquad \nabla h_j(x^0), \quad j \in I^*, \tag{4.42}$$

линейно независимы.

Следующая теорема описывает связь между этими условиями.

Теорема 4.19. *Имеют место соотношения*

$$R_1^* \Leftarrow R_3^* \Longleftrightarrow R_2^* \Leftarrow R_4^*.$$

Доказательство. 1. $R_1^* \Leftarrow R_3^*$. Доказательство аналогично доказательству теоремы 4.11. При этом необходимо лишь заменить $I_0 \setminus I_1$ на I_0, I_1 на I^*, G_1 на G^*, $T_1(x^0)$ на $T^*(x^0)$, $K_1(x^0)$ на $K^*(x^0)$.

2. $R_2^* \Longleftrightarrow R_3^*$. Доказательство следует из теоремы Моцкина об альтернативе (теорема 2.106) с учетом линейной независимости $\nabla h_j(x^0)$ для $j \in I^*$.

3. $R_2^* \Leftarrow R_4^*$ следует непосредственно из определения линейной независимости системы векторов.

Для формулировки необходимых условий второго порядка введем следующие множества:

$$G_1^* = G^* \cap \{x\colon\ g_i(x) = 0 \quad \forall i \in I_1\}, \quad I_1 = \{i \in I_0\colon\ y_i^0 > 0\},$$

$$K_1^*(x^0) = \{z\colon\ \nabla g_i(x^0)'z \leqslant 0 \quad \forall i \in I_0 \setminus I_1;$$

$$\nabla g_i(x^0)'z = 0 \quad \forall i \in I_1;\ \nabla h_j(x^0)'z = 0 \quad \forall j \in I^*\},$$

$T_1^*(x^0)$ – касательный конус к G_1^* в точке x^0.

Кроме того, нам понадобится условие

B_1^*: функции $g\colon \mathbf{R}^n \to \mathbf{R}^m$, $h\colon \mathbf{R}^n \to \mathbf{R}^p$ дифференцируемы в точке $x^0 \in G^*$ и справедливо равенство

$$T_1^*(x^0) = K_1^*(x^0).$$

Применяя введенные обозначения, мы получаем следующее утверждение относительно задач P_1^* и P_3^*.

Теорема 4.20. *Пусть* $f\colon \mathbf{R}^n \to \mathbf{R}$, $g\colon \mathbf{R}^n \to \mathbf{R}^m$, $h\colon \mathbf{R}^n \to \mathbf{R}^p$ *— дважды дифференцируемые в* $x^0 \in G^*$ *функции,* x^0 *– точка локального минимума* f *на множестве* G^*, $(x^0, y^0, v^0)'$ *– решение задачи* P_3^*. *Предположим, что в точке* x^0 *выполнены условие регулярности* R_1^* *и условие* B_1^*. *Тогда*

$$z'\nabla_x^2 L^*(x^0, y^0, v^0)z \geqslant 0 \qquad \forall\, z \in K_1^*(x^0).$$

Доказательство использует теорему 4.4 и аналогично доказательству теоремы 4.10.

Сформулируем и здесь условие, достаточное для выполнения B_1^*:

B_2^*: функции $g\colon \mathbf{R}^n \to \mathbf{R}^m$, $h\colon \mathbf{R}^n \to \mathbf{R}^p$ дифференцируемы в точке $x^0 \in G^*$, существует вектор $z \in \mathbf{R}^n$ такой, что

$$\nabla g_i(x^0)'z < 0 \qquad \forall\, i \in I_0 \setminus I_1, \qquad \nabla g_i(x^0)'z = 0 \qquad \forall\, i \in I_1,$$

$$\nabla h_j(x^0)'z = 0 \qquad \forall\, j \in I^*,$$

градиенты $\nabla g_i(x^0)$, $i \in I_1$, $\nabla h_j(x^0)$, $j \in I^*$, линейно независимы.

Теорема 4.21. *Из условия* B_2^* *следует условие* B_1^*.

Доказательство аналогично доказательству теоремы 4.11.

Замечание. Поскольку имеют место импликации $R_4^* \Rightarrow B_2^*$ и $R_4^* \Rightarrow R_1^*$, то в теореме 4.20 условия B_1^* и R_1^* могут быть заменены условием R_4^*.

С учетом теоремы 4.5 получаем следующее утверждение.

Теорема 4.22. *Пусть функции* $f\colon \mathbf{R}^n \to \mathbf{R}$, $g\colon \mathbf{R}^n \to \mathbf{R}^m$, $h\colon \mathbf{R}^n \to \mathbf{R}^p$ *дважды дифференцируемы в точке* $x^0 \in G^*$ *и* $(x^0, y^0, v^0)' \in \mathbf{R}^n \times \mathbf{R}_+^m \times \times \mathbf{R}^p$ *– решение задачи* P_3^*, *удовлетворяющее соотношению*

$$z'\nabla_x^2 L^*(x^0, y^0, v^0)z > 0 \tag{4.43}$$

для всех $z \in T^*(x^0) \cap \{z\colon \nabla g_i(x^0)'z = 0 \ \ \forall\, i \in I_1\}$, $z \neq 0$. *Тогда* x^0 *является точкой собственного локального минимума функции* f *на множестве* G^*.

Замечание. Как и в теореме 4.12, условие (4.43) можно усилить. В силу включения $T_1^*(x^0) \subseteq K_1^*(x^0)$ новым условием будет

$$z'\nabla_x^2 L^*(x^0, y^0, v^0)z > 0 \qquad \forall\, z \in K_1^*(x^0), \quad z \neq 0.$$

Предположим (отходя от прежнего определения), что множество I пусто. Тогда из теорем 4.18, 4.20 и 4.22 формально вытекает классическая теория множителей Лагранжа. При $I^* = \phi$ имеем результаты п. 4.2.2.

4.3. Глобальная теория множителей Лагранжа

Рассмотрим задачу

$$P_X\colon \min \{f(x)\colon x \in G_X\},$$

где $G_X = \{x \in \mathbf{R}^n\colon g(x) \leqslant 0,\ x \in X\}$, $X \subseteq \mathbf{R}^n$, $X \neq \phi$, $f\colon X \to \mathbf{R}$, $g\colon X \to \mathbf{R}^m$.

Если положить $X = \mathbf{R}^n$ то из задачи P_X получим как частный случай задачу P_1. В дальнейшем основное внимание мы уделим формулировке критериев оптимальности в форме теорем о седловой точке. Центральная теорема из п. 4.3 будет позднее приведена еще раз в несколько ином виде как сильная теорема двойственности в 5.4. Там будет дана также геометрическая интерпретация результатов этого параграфа. Вообще, глобальную теорию множителей Лагранжа содержательно нельзя отделять от теорем двойственности.

Будем говорить, что функция $S\colon X \times Y \to \mathbf{R}$, $X \subseteq \mathbf{R}^n$, $Y \subseteq \mathbf{R}^m$, имеет в $(x^0, y^0)' \in X \times Y$ *седловую точку на множестве* $X \times Y$, если

$$S(x^0, y) \leqslant S(x^0, y^0) \leqslant S(x, y^0)$$

для всех $(x, y)' \in X \times Y$.

Для формулировки глобальных критериев оптимальности рассмотрим следующую задачу.

Задача о седловой точке $S_{\eta,X}$: найти точки $x^0 \in X$ и $(\eta^0, y^0)' \in \mathbf{R}_+^{m+1}$, $(\eta^0, y^0)' \neq 0$, такие, чтобы для функции $L_\eta(x, y) = \eta f(x) + y' g(x)$, $(x, y)' \in X \times \mathbf{R}_+^m$ выполнялось соотношение

$$L_{\eta^0}(x^0, y) \leqslant L_{\eta^0}(x^0, y^0) \leqslant L_{\eta^0}(x, y^0)$$

для всех $(x, y)' \in X \times \mathbf{R}_+^m$.

Можно привести следующий необходимый критерий оптимальности для задачи P_X.

Теорема 4.23. *Пусть $X \subseteq \mathbf{R}^n$ – выпуклое множество, $f\colon X \to \mathbf{R}$ и $g\colon X \to \mathbf{R}^m$ – выпуклые функции. Если x^0 – решение задачи P_X, то существуют $\eta^0 \in \mathbf{R}_+$ и $y^0 \in \mathbf{R}_+^m$ такие, что точка $(x^0, \eta^0, y^0)' \in X \times \mathbf{R}_+^{m+1}$ является решением задачи $S_{\eta,X}$.*

Доказательство. Пусть x^0 – решение задачи P_X. Тогда система

$$f(x) - f(x^0) < 0, \quad g(x) \leqslant 0, \quad x \in X,$$

несовместна. По теореме 2.103 найдется вектор $(\eta^0, y^0)' \in \mathbf{R}_+^{m+1}$, $(\eta^0, y^0)' \neq 0$, такой, что ,

$$L_{\eta^0}(x, y^0) = \eta^0 f(x) + y^{0\prime} g(x) \geqslant \eta^0 f(x^0)$$

для всех $x \in X$. Отсюда в силу $g(x^0) \leqslant 0$ имеем

$$L_{\eta^0}(x, y^0) = \eta^0 f(x) + y^{0\prime} g(x) \geqslant \eta^0 f(x^0) + y' g(x^0) = L_{\eta^0}(x^0, y) \qquad (4.44)$$

для всех $(x, y)' \in X \times \mathbf{R}_+^m$. Полагая в (4.44) $y = y^0$, получим

$$\eta^0 f(x) + y^{0\prime} g(x) \geqslant \eta^0 f(x^0) + y^{0\prime} g(x^0) \quad \forall x \in X; \qquad (4.45)$$

полагая в (4.44) $x = x^0$, получим

$$\eta^0 f(x^0) + y^{0\prime} g(x^0) \geqslant \eta^0 f(x^0) + y' g(x^0) \quad \forall y \in \mathbf{R}_+^m. \qquad (4.46)$$

Неравенства (4.45) и (4.46) доказывают теорему.

Утверждение теоремы 4.23 имеет сходство с теоремой 4.8 в следующем смысле: для заданной области $G_X = \{x: g(x) \leqslant 0, x \in X\}$ приведенное в теореме необходимое условие может выполняться для произвольных целевых функций, поскольку возможен случай $\eta^0 = 0$. Поэтому, как и в локальной теории множителей Лагранжа, будем искать условие регулярности, которое обеспечивает неравенство $\eta^0 > 0$. Для этого рассмотрим следующую задачу.

Задача о седловой точке S_X: найти седловую точку $(x^0, y^0)' \in X \times \mathbf{R}_+^m$ функции Лагранжа

$$L(x, y) = f(x) + y'g(x)$$

на множестве $X \times \mathbf{R}_+^m$.

Применим условие регулярности R_S^X: множество $X \subseteq \mathbf{R}^n$ выпукло, функции $g_i: X \to \mathbf{R}$, $i \in I$, выпуклы и существует точка $\bar{x} \in \operatorname{ri} X$ такая, что

$$g_i(\bar{x}) < 0, \quad i \in I^N, \quad g_i(\bar{x}) \leqslant 0, \quad i \in I^L.$$

Теорема 4.24. *Пусть $X \subseteq \mathbf{R}^n$ – выпуклое множество, $f: X \to \mathbf{R}$, $g_i: X \to \mathbf{R}$, $i \in I$, – выпуклые функции и выполнено условие регулярности R_S^X. Если $x^0 \in X$ – решение задачи P_X, то существует вектор $y^0 \in \mathbf{R}_+^m$ такой, что $(x^0, y^0)' \in X \times \mathbf{R}_+^m$ будет решением задачи S_X.*

Доказательство. По предположению система

$$f(x) - f(x^0) < 0, \quad g(x) \leqslant 0, \quad x \in X,$$

не имеет решений. Напротив, в силу R_S^X система

$$g_{I^N}(x) < 0, \quad g_{I^L}(x) \leqslant 0, \quad x \in \operatorname{ri} X,$$

совместна. По теореме 2.104 найдется вектор $(\eta^0, y^0)' \in \mathbf{R}_+^{m+1}$, $\eta^0 \neq 0$, со свойством

$$\eta^0(f(x) - f(x^0)) + y^{0\prime}g(x) \geqslant 0 \quad \forall x \in X.$$

Без ограничения общности можем считать $\eta^0 = 1$, т.е.

$$L(x, y^0) = f(x) + y^{0\prime}g(x) \geqslant f(x^0) \quad \forall x \in X.$$

В силу того, что $g(x^0) \leqslant 0$, имеем

$$L(x^0, y) = f(x^0) + y'g(x^0) \leqslant f(x^0) \quad \forall y \in \mathbf{R}_+^m,$$

и потому

$$L(x^0, y^0) = f(x^0) + y^{0\prime}g(x^0) = f(x^0).$$

Отсюда следует, что $(x^0, y^0)'$ – седловая точка функции $L(x, y)$ на множестве $X \times \mathbf{R}_+^m$.

Сформулируем теперь достаточные условия для существования решения задачи P_X. Оказывается, что без дополнительных предположений о задаче P_X справедлива теорема, обратная к теореме 4.24.

Теорема 4.25. *Если $(x^0, y^0)' \in X \times \mathbf{R}_+^m$ – решение задачи S_X, то x^0 – решение P_X.*

Доказательство. Из неравенства $L(x^0, y) \leqslant L(x^0, y^0)$, справедливого для всех $y \in \mathbf{R}_+^m$, следует

$$(y - y^0)'g(x^0) \leqslant 0 \qquad \forall y \in \mathbf{R}_+^m. \tag{4.47}$$

Если в (4.47) положить $y = y^0 + y^1$, $y^1 \in \mathbf{R}_+^m$, то получим $y^{1\prime}g(x^0) \leqslant 0$ $\forall y^1 \in \mathbf{R}_+^m$. Допуская, что $g_{i_0}(x^0) > 0$ для $i_0 \in I$, приходим к противоречию при $y_{i_0}^1 > 0$ и $y_i^1 = 0$, $i \neq i_0$. Поэтому $g(x^0) \leqslant 0$.

Из (4.47) при $y = 0$ вытекает неравенство $y^{0\prime}g(x^0) \geqslant 0$. В силу того, что $g(x^0) \leqslant 0$ и $y^0 \geqslant 0$, получаем $y^{0\prime}g(x^0) = 0$. Из $L(x^0, y^0) \leqslant L(x, y^0)$ $\forall x \in X$ следует $f(x^0) \leqslant f(x) + y^{0\prime}g(x)$ $\forall x \in X$, и тем самым для $x \in G_X$ сразу же

$$f(x^0) \leqslant f(x) + y^{0\prime}g(x) \leqslant f(x).$$

Таким образом, с помощью теоремы 4.25 без каких-либо предположений о целевой функции и функциях ограничений из существования седловой точки функции L на множестве $X \times \mathbf{R}_+^m$ можно сделать вывод о существовании решения задачи P_X. Возникает вопрос, не является ли для $X = \mathbf{R}^n$ разрешимость задачи P_3 при некоторых условиях достаточной также для разрешимости $P_X = P_1$. Действительно, такое заключение удается сделать для псевдовыпуклой целевой функции f и дифференцируемых квазивыпуклых функций ограничений g_i, $i \in I$.

Теорема 4.26. *Пусть* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и* $g_i\colon \mathbf{R}^n \to \mathbf{R}$, $i \in I$, *– дифференцируемые в точке* x^0 *функции,* f *псевдовыпукла, а* g_i, $i \in I$, *квазивыпуклы. Если* $(x^0, y^0)' \in \mathbf{R}^n \times \mathbf{R}_+^m$ *– решение задачи* P_3, *то* x^0 *– решение* P_1.

Доказательство. 1. Пусть $g(x^0) < 0$. Тогда $y^0 = 0$ и, поскольку $\nabla_x L(x^0, y^0) = 0$, следует $\nabla f(x^0) = 0$. По теореме 2.102 функция f достигает в точке x^0 глобального минимума.

2. Для $I_0 \neq \phi$ имеем $g_i(x) \leqslant g_i(x^0)$ $\forall x \in G$ $\forall i \in I_0$. Из квазивыпуклости g_i следует

$$\nabla g_i(x^0)'(x - x^0) \leqslant 0 \qquad \forall x \in G \qquad \forall i \in I_0.$$

Учитывая, что $y^0 \geqslant 0$ и $y_i^0 = 0$ для $i \in I \setminus I_0$, имеем

$$y_i^0 \nabla g_i(x^0)'(x - x^0) \leqslant 0 \qquad \forall x \in G, \qquad \forall i \in I. \tag{4.48}$$

Так как $\nabla_x L(x^0, y^0) = 0$, то из (4.48) получаем

$$\nabla f(x^0)'(x - x^0) \geqslant 0 \qquad \forall x \in G.$$

Отсюда в силу псевдовыпуклости f согласно теореме 2.95 получаем

$$f(x) \geqslant f(x^0) \qquad \forall x \in G.$$

Заметим, что в предположениях теоремы 4.26 не может быть сделан вывод о существовании седловой точки функции L на множестве $\mathbf{R}^n \times \mathbf{R}_+^m$. Это показывает следующий пример.

Пример 4.9. В задаче

$$P_1\colon \min\{f(x) = -e^{-x^2}\colon g_1(x) = 1 - x \leqslant 0,\ x \in \mathbf{R}\}$$

целевая функция f – псевдовыпуклая, а функция ограничений g_1 – аффинная. Кроме того, для $(x^0, y^0)' = (1, 2/e)'$ имеем

$$\nabla_x L(x^0, y^0) = 2x^0 e^{-x^2} - y^0 = 0, \qquad y^0 g(x^0) = 0, \qquad g(x^0) < 0, \qquad y^0 = 2/e > 0.$$

Таким образом, все предположения теоремы 4.26 выполнены. Следовательно, $x^0 = 1$ – решение задачи P_1. С другой стороны, $(x^0, y^0)'$ не является седловой точкой функции L на $\mathbf{R} \times \mathbf{R}_+$, поскольку не для всех $x \in \mathbf{R}$ справедливо

$$-1/e = L(x^0, y^0) \leqslant L(x, y^0) = -e^{-x^2} + 2(1-x)/e.$$

Следует вспомнить о том, что неотрицательная линейная комбинация квазивыпуклых (или псевдовыпуклых) функций не обязательно дает снова квазивыпуклую (псевдовыпуклую) функцию. Примером этого как раз и является функция $L(x, y^0) = -e^{-x^2} + 2(1-x)/e$.

Если усилить предположения теоремы 4.26, потребовав выпуклости функций f и g_i, $i \in I$, то из разрешимости задачи P_3 все же можно сделать вывод о разрешимости S_X (при $X = \mathbf{R}^n$).

4.4. Критерии седловой точки для функции Лагранжа

Т е о р е м а 4.27. *Пусть* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и* $g_i\colon \mathbf{R}^n \to \mathbf{R}$, $i \in I$, – *выпуклые дифференцируемые в точке* x^0 *функции. Если* $(x^0, y^0)' \in \mathbf{R}^n \times \mathbf{R}^m_+$ – *решение задачи* P_3, *то* $(x^0, y^0)'$ – *решение* S_X *при* $X = \mathbf{R}^n$.

Д о к а з а т е л ь с т в о. С учетом выпуклости $L(x, y^0)$ по x имеем

$$L(x, y^0) - L(x^0, y^0) \geqslant \nabla_x L(x^0, y^0)'(x - x^0) = 0 \quad \forall x \in \mathbf{R}^n.$$

Далее, из соотношений $y^{0\prime} g(x^0) = 0$, $g(x^0) \leqslant 0$, $y^0 \geqslant 0$ следует неравенство

$$L(x^0, y^0) \geqslant L(x^0, y) \quad \forall y \in \mathbf{R}^m_+.$$

Т е о р е м а 4.28. *Пусть функции* $f\colon \mathbf{R}^n \to \mathbf{R}$ *и* $g_i\colon \mathbf{R}^n \to \mathbf{R}$, $i \in I$, *дифференцируемы в точке* x^0. *Если* $(x^0, y^0)' \in \mathbf{R}^n \times \mathbf{R}^m_+$ – *решение задачи* S_X *при* $X = \mathbf{R}^n$, *то* $(x^0, y^0)'$ – *решение* P_3.

Д о к а з а т е л ь с т в о. Поскольку $(x^0, y^0)'$ – *седловая точка* L *на* $\mathbf{R}^n \times \mathbf{R}^m_+$, то $g(x^0) \leqslant 0$ и $y^{0\prime} g(x^0) = 0$ (см. доказательство теоремы 4.25).

Из неравенства $L(x, y^0) - L(x^0, y^0) \geqslant 0$ $\forall x \in \mathbf{R}^n$ при $x = x^0 + tz$, $t \in \mathbf{R}$, получаем

$$\lim_{t \to 0} \frac{L(x^0 + tz, y^0) - L(x^0, y^0)}{t} = \nabla_x L(x^0, y^0)' z \geqslant 0 \quad \forall z \in \mathbf{R}^n,$$

откуда следует $\nabla_x L(x^0, y^0) = 0$.

Последние две теоремы связывают утверждения локальной и глобальной теорий множителей Лагранжа. В § 4.5 мы дадим обзор структуры связей между рассмотренными задачами для случая $X = \mathbf{R}^n$.

В заключение мы приведем одно утверждение о существовании седловой точки функции, указывающее на связь с теоремами о минимаксе, известными, например, из теории игр (см. также § 5.10).

Для произвольной функции $S\colon X \times Y \to \mathbf{R}$, где $X \subseteq \mathbf{R}^n$, $Y \subseteq \mathbf{R}^m$, утверждение о выполнимости соотношения

$$\min_{x \in X} \max_{y \in Y} S(x, y) = \max_{y \in Y} \min_{x \in X} S(x, y)$$

называют *теоремой о минимаксе.*

Т е о р е м а 4.29. *Функция* S: $X \times Y \to \mathbf{R}$, $X \subseteq \mathbf{R}^n$, $Y \subseteq \mathbf{R}^m$, *тогда и только тогда имеет в* $(x^0, y^0)' \in X \times Y$ *седловую точку на множестве* $X \times Y$, *когда*

$$S(x^0, y^0) = \min_{x \in X} \max_{y \in Y} S(x, y) = \max_{y \in Y} \min_{x \in X} S(x, y). \tag{4.49}$$

Д о к а з а т е л ь с т в о. Имеют место неравенства

$$\sup_{y \in Y} S(x, y) \geqslant S(x, y) \geqslant \inf_{x \in X} S(x, y) \quad \forall (x, y)' \in X \times Y,$$

$$\inf_{x \in X} \sup_{y \in Y} S(x, y) \geqslant \sup_{y \in Y} \inf_{x \in X} S(x, y).$$

Поэтому для произвольной точки $(\bar{x}, \bar{y})' \in X \times Y$ получаем

$$\sup_{y \in Y} S(\bar{x}, y) \geqslant \inf_{x \in X} \sup_{y \in Y} S(x, y) = \sup_{y \in Y} \inf_{x \in X} S(x, y) \geqslant \inf_{x \in X} S(x, \bar{y}). \tag{4.50}$$

Кроме того, очевидно, что функция S тогда и только тогда имеет в $(x^0, y^0)' \in X \times Y$ седловую точку на $X \times Y$, когда

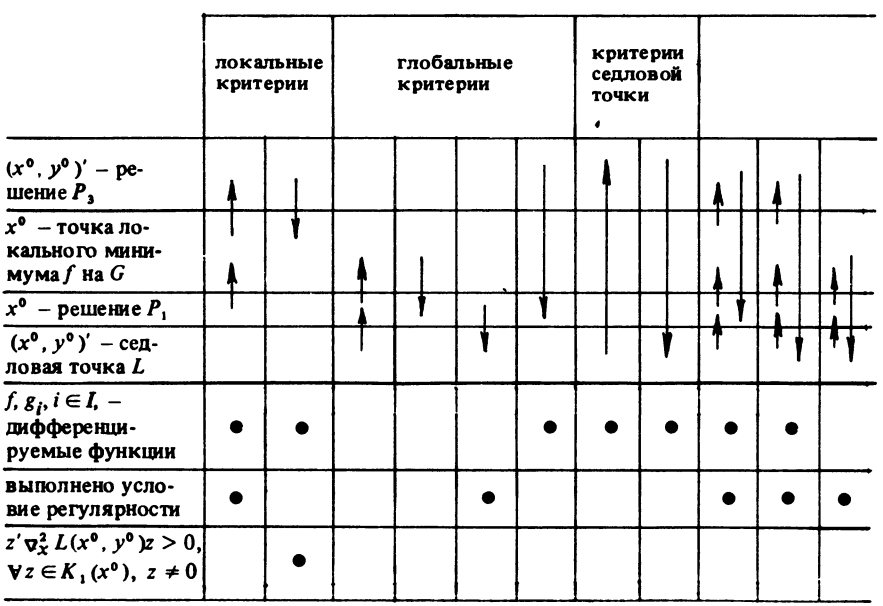
$$S(x^0, y^0) = \max_{y \in Y} S(x^0, y) = \min_{x \in X} S(x, y^0). \tag{4.51}$$

1. Если функция S имеет в $(x^0, y^0)'$ седловую точку, то соотношение (4.49) тотчас же следует из (4.51) и (4.50).

2. Обратно, из (4.49) сразу же вытекает (4.51), и, тем самым, существование седловой точки $(x^0, y^0)'$ функции S на множестве $X \times Y$.

4.5. Итоговый обзор

Между рассмотренными выше задачами существуют следующие взаимосвязи:



	локальные критерии		глобальные критерии				критерии седловой точки				
f псевдовыпукла, g_i, $i \in I$, квазивыпуклы						●			●		
f, g_i, $i \in I$, выпуклы				●	●			●		●	●

Данная таблица суммирует результаты главы 4. В ее нижней части приведены условия, при выполнении которых справедливы импликации, указанные в верхней части (например, первый столбец соответствует теореме 4.9).

Из таблицы следует, что в случае выпуклых дифференцируемых функций при выполнении условия регулярности разрешимость одной из задач, включенных в обзор, влечет разрешимость всех остальных.

5.1. Двойственные задачи математического программирования

Мы обсудим в дальнейшем два пути для вывода теорем двойственности. Первый подход, который излагается подробно, опирается на теорию сопряженных функций и одновременно проясняет связи с теорией Куна – Таккера и анализом чувствительности *). Второй подход обсуждается нами не столь полно и основывается на применении теорем о минимаксе.

Рассмотрим *прямую задачу*

$$P_X\colon \min\{f(x)\colon x\in G\},\qquad G=\{x\colon g(x)\leqslant 0,\ x\in X\},$$

где $X\subseteq \mathbf{R}^n$ – непустое множество, а f, g – отображения вида $f\colon X\to\mathbf{R}$, $g\colon X\to\mathbf{R}^m$.

Применяя функцию Лагранжа L, задаче P_X поставим в соответствие *двойственную задачу*

$$D_X\colon \max\{\varphi(y)\colon y\in\mathbf{R}^m_+\},$$

где

$$\varphi(y)=\inf_{x\in X} L(x,y)=\inf_{x\in X}\{f(x)+y'g(x)\},\qquad y\in\mathbf{R}^m_+.$$

Очевидно, что φ – отображение вида $\varphi\colon \mathbf{R}^m_+\to\mathbf{R}\cup\{-\infty\}$.

Исследуем соотношения между решениями пары двойственных задач P_X и D_X.

Обозначим через α и β оптимальные значения прямой и двойственной задач соответственно, т.е.

$$\alpha=\inf_{x\in G} f(x),\qquad \beta=\sup_{y\in\mathbf{R}^m_+}\varphi(y).$$

Если при этом $G=\phi$, то условимся, что $\alpha=+\infty$.

Непосредственно из определения задачи D_X для произвольных функций $f\colon X\to\mathbf{R}$ и $g\colon X\to\mathbf{R}^m$ вытекает *слабая теорема двойственности*.

Т е о р е м а 5.1. *Для любых* $(\bar x,\bar y)'\in G\times\mathbf{R}^m_+$ *справедливо неравенство* $f(\bar x)\geqslant\varphi(\bar y)$.

В частности,

$$\alpha\geqslant\beta. \tag{5.1}$$

Д о к а з а т е л ь с т в о. Для произвольной пары точек $(\bar x,\bar y)$, допустимых соответственно в прямой и двойственной задачах, имеем

$$\varphi(\bar y)=\inf_{x\in X}\{f(x)+\bar y'g(x)\}\leqslant f(\bar x)+\bar y'g(\bar x)\leqslant f(\bar x).$$

*) См. также [32, 55, 74].

Отсюда и следует утверждение теоремы.

Главная цель теории двойственности состоит в том, чтобы охарактеризовать задачи, для которых в (5.1) имеет место знак равенства. Кроме того, речь пойдет о существовании решений задач P_X и D_X.

Если в (5.1) выполняется строгое неравенство, то говорят, что для задач P_X и D_X имеется *разрыв в двойственности.*

Из теоремы 5.1 видно, что совокупность значений двойственной целевой функции φ на $\mathbf{R}_+^m$ задает нижние границы для значений прямой целевой функции f на множестве G. Это особенно важно при численном анализе конкретной задачи математического программирования. Простые примеры показывают, что практическое значение полученных таким образом оценок зависит от наличия разрывов в двойственности.

Пример 5.1. Для $X = [-1, \infty) \subset \mathbf{R}$; $f(x) = -x$; $g(x) = 1$ получаем $\varphi(y) = -\infty$ $\forall y \in \mathbf{R}_+$. Тем самым имеем

$$+\infty = \alpha = \inf_{x \in G} (-x) > \beta = \sup_{y \in \mathbf{R}_+} \varphi(y) = -\infty.$$

Но и конечность β не позволяет сделать вывод о "качестве" упомянутой выше оценки.

Пример 5.2. Для $X = (0, 1) \subset \mathbf{R}$; $f(x) = x$; $g(x) = x$ получаем

$$\varphi(y) = \inf_{x \in X} \{(1 + y)x\} = 0, \quad y \in \mathbf{R}_+,$$

$$+\infty = \alpha = \inf_{x \in X \cap -\mathbf{R}_+} x > \beta = \sup_{y \in \mathbf{R}_+} \varphi(y) = \varphi(0) = 0.$$

Менее тривиальные примеры можно найти в § 5.5 (примеры 5.8 и 5.10).

5.2. Свойства функции чувствительности и двойственной целевой функции

В целях построения теории двойственности для задач P_X и D_X на основе теории сопряженных функций введем понятие функции чувствительности прямой задачи.

Функцией чувствительности задачи P_X назовем отображение q: $\mathbf{R}^m \to \overline{\mathbf{R}}$ вида

$$q(v) = \inf_{x \in X} \{f(x): g(x) \leqslant v\}, \quad v \in \mathbf{R}^m. \tag{5.2}$$

Положим $G_v = \{x \in X: g(x) \leqslant v\}$, $v \in \mathbf{R}^m$. Справедливы соотношения

$$\alpha = q(0), \quad X = \bigcup_{v \in \mathbf{R}^m} G_v, \quad G = G_0,$$

$$\operatorname{dom} q = \{v \in \mathbf{R}^m: q(v) < +\infty\} = \{v \in \mathbf{R}^m: G_v \neq \phi\}. \tag{5.3}$$

Без предположений о выпуклости функций f и g имеет место следующая теорема.

Теорема 5.2. *Пусть q – функция чувствительности произвольной задачи P_X. Тогда*:

1) *q монотонно убывает, т.е. из $v^1 \leqslant v^2$ следует $q(v^1) \geqslant q(v^2)$;*

2) $\operatorname{int} \operatorname{dom} q \neq \phi$.

Доказательство. 1. Из $v^1 \leqslant v^2$ следует $\inf\limits_{x \in G_{v^1}} f(x) \geqslant \inf\limits_{x \in G_{v^2}} f(x)$.

2. По предположению $X \neq \phi$. Пусть $x^0 \in X$. Положим $v^0 = g(x^0)$. Тогда для всех $v \geqslant v^0$ имеем

$$G_v \neq \phi, \quad q(v) = \inf_{x \in G_v} f(x) \leqslant f(x^0) < \infty,$$

и, следовательно, $\operatorname{int} \operatorname{dom} q \neq \phi$.

В теоремах двойственности большое значение имеет выпуклость функции чувствительности. Одно достаточное (но не необходимое) условие для этого сформулировано в следующей теореме.

Теорема 5.3. *Пусть $X \subseteq \mathbf{R}^n$ – выпуклое множество, а $f\colon X \to \mathbf{R}$, $g\colon X \to \mathbf{R}^m$ – выпуклые функции. Тогда функция чувствительности q задачи P_X выпукла.*

Доказательство. Прежде всего вспомним, что

$$G_y = \{x \in X\colon g(x) \leqslant y\}, \quad y \in \mathbf{R}^m. \tag{5.4}$$

Пусть $y^1, y^2 \in \mathbf{R}^m$, $q(y^i) \in \mathbf{R}$, $\mu_i \geqslant 0$, $\mu_1 + \mu_2 = 1$. Тогда

$$q(y) = \inf_{x \in G_y} f(x), \quad q(\mu_1 y^1 + \mu_2 y^2) = \inf_{x \in G_{\mu_1 y^1 + \mu_2 y^2}} f(x).$$

Если $x^1 \in G_{y^1}$, $x^2 \in G_{y^2}$, то в силу неравенства

$$g(\mu_1 x^1 + \mu_2 x^2) \leqslant \mu_1 g(x^1) + \mu_2 g(x^2) \leqslant \mu_1 y^1 + \mu_2 y^2$$

из (5.4) следует $\mu_1 x^1 + \mu_2 x^2 \in G_{\mu_1 y^1 + \mu_2 y^2}$.

Тем самым получаем

$$\inf_{x \in G_{\mu_1 y^1 + \mu_2 y^2}} f(x) \leqslant \inf_{(x^1, x^2)' \in G_{y^1} \times G_{y^2}} f(\mu_1 x^1 + \mu_2 x^2) \leqslant$$

$$\leqslant \inf_{(x^1, x^2)' \in G_{y^1} \times G_{y^2}} \{\mu_1 f(x^1) + \mu_2 f(x^2)\} = \mu_1 \inf_{x^1 \in G_{y^1}} f(x^1) +$$

$$+ \mu_2 \inf_{x^2 \in G_{y^2}} f(x^2) = \mu_1 q(y^1) + \mu_2 q(y^2).$$

Между функцией чувствительности q прямой задачи и двойственной целевой функцией φ имеется следующая связь.

Теорема 5.4. *Если q – функция чувствительности задачи P_X, то*

$$\varphi(y) = -q^c(-y) = \inf_{v \in \mathbf{R}^m} \{q(v) + y'v\}, \quad y \in \mathbf{R}^m_+;$$

в частности, отсюда вытекает, что φ – вогнутая функция.

Доказательство. 1. По определению сопряженной функции имеем

$$-q^c(-y) = \inf_{v \in \mathbf{R}^m} \{q(v) + y'v\}.$$

Положим

$$F(x, v) = \begin{cases} f(x), & \text{если } g(x) \leqslant v, \\ +\infty & \text{в противном случае,} \end{cases}$$

и так как $q(v) = \inf\limits_{x \in X} F(x, v)$, получим

$$-q^c(-y) = \inf_{v \in \mathbf{R}^m} \inf_{x \in X} \{F(x, v) + y'v\}.$$

Поскольку переменные x и v независимы друг от друга, то для $y \in \mathbf{R}^m_+$ имеем

$$-q^c(-y) = \inf_{x \in G_v} \inf_{v \in \operatorname{dom} q} \{f(x) + y'v\} =$$

$$= \inf_{x \in X} \{f(x) + y'g(x)\} = \varphi(y), \qquad y \in \mathbf{R}^m_+.$$

2. Так как функция q^c выпукла, то φ вогнута. Кроме того, множество $\{y\colon \varphi(y) > -\infty\}$ выпукло.

С учетом теоремы 5.4, а также монотонности функции чувствительности q задачи P_X и D_X могут быть представлены в виде

$$P_X\colon \min\{q(v)\colon v \in -\mathbf{R}^m_+\}, \qquad D_X\colon \max\{-q^c(-y)\colon y \in \mathbf{R}^m_+\}.$$

В силу того, что $\operatorname{dom} q \neq \phi$, можно записать

$$\sup_{y \in \mathbf{R}^m_+} \varphi(y) = \sup_{-y \in \operatorname{dom} q^c} \{-q^c(-y)\} =$$

$$= \sup_{-y \in \operatorname{dom} q^c} \{0'(-y) - q^c(-y)\} = q^{cc}(0). \tag{5.5}$$

Поэтому утверждение теоремы 5.1 можно сформулировать так:

$$q(0) \geqslant q^{cc}(0). \tag{5.6}$$

Это соотношение нам уже известно из § 3.2.

Предыдущие рассуждения позволяют наглядно интерпретировать теорию двойственности, а также глобальную теорию множителей Лагранжа (см. § 4.4). Соотношение $\varphi(y) = \inf\limits_{v \in \mathbf{R}^m} \{q(v) + y'v\}$ означает, что невертикаль-

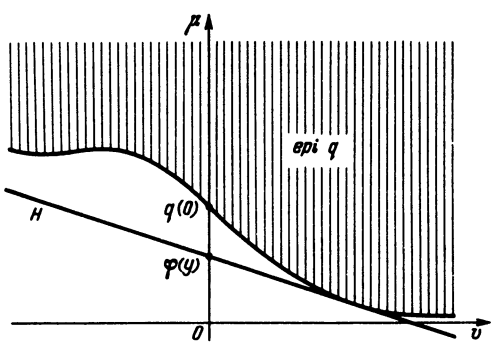


Рис. 5.1.

ная гиперплоскость

$$H = \{(v, \mu)' \in \mathbf{R}^{m+1}: \ \mu + y'v = \varphi(y)\}$$

является опорной гиперплоскостью к множеству epi$q$ и что (см. рис. 5.1)

$$\text{epi}\, q \subseteq H_+, \qquad H_+ = \{(v, \mu)' \in \mathbf{R}^{m+1}: \ \mu + y'v \geqslant \varphi(y)\}.$$

Вектор $y \in \mathbf{R}_+^m$ такой, что H является невертикальной гиперплоскостью, опорной к epi$q$ в точке $(0, q(0))'$, существует не всегда. Теорема 5.1 говорит лишь о том, что

$$\varphi(y) \leqslant q(0) \qquad \forall y \in \mathbf{R}_+^m, \tag{5.7}$$

а это означает, что значение координаты μ точки пересечения H с осью ординат не превосходит $q(0)$. Если $\sup\limits_{y \in \mathbf{R}_+^m} \varphi(y) < q(0)$, то для задач P_X и D_X имеется разрыв в двойственности (см. рис. 5.1).

В следующем параграфе мы укажем некоторые соотношения эквивалентности с тем, чтобы сформулировать потом достаточные условия для выполнения равенства в (5.1) и (5.6).

5.3. Другие утверждения о функции чувствительности

Принимая во внимание равенства (5.5), тотчас же получаем следующую теорему.

Теорема 5.5. *Для выполнения соотношения*

$$\alpha = \inf_{x \in G} f(x) = \sup_{y \in \mathbf{R}_+^m} \varphi(y) = \beta$$

необходимо и достаточно, чтобы

$$q(0) = q^{cc}(0).$$

Теорема 5.5 включает также случаи $\alpha = \beta = +\infty$ и $\alpha = \beta = -\infty$. Покажем это на следующих примерах.

Пример 5.3. Для $X = (-\infty, 1] \subset \mathbf{R}$, $f(x) = x + 1$, $g(x) = x$ находим $q(v) \equiv -\infty$, $\varphi(y) = -\infty \ \forall y \in \mathbf{R}_+$. Поэтому $\alpha = \beta = -\infty$.

Пример 5.4. Для $X = [1, 2] \subset \mathbf{R}$, $f(x) = x + 1$, $g(x) = x$ находим

$$q(v) = \begin{cases} 2, & \text{если } v \in [1, \infty), \\ +\infty & \text{в противном случае,} \end{cases} \qquad \varphi(y) = 2 + y, \quad y \in \mathbf{R}_+.$$

Отсюда $\alpha = \beta = +\infty$.

Теорема 5.5 следующим образом характеризует функции чувствительности, которые обеспечивают равенство в (5.1) или (5.7): точка $(0, q(0))' \in \in \mathbf{R}^{m+1}$ должна принадлежать графику поточечно наибольшей полунепрерывной снизу выпуклой функции $\tilde{q} = q^{cc}$, которая расположена ниже q (см. в этой связи теорему 3.16).

Следующие примеры показывают, что из существования решения одной из задач P_X или D_X нельзя сделать вывод о существовании решения другой из них (также и при $\alpha = \beta$).

Пример 5.5. Для $X = \mathbf{R}$, $f(x) = e^{-x}$, $g(x) = -x$ находим

$$q(0) = \inf_{x \in \mathbf{R}_+} e^{-x} = 0, \quad \varphi(y) = \inf_{x \in \mathbf{R}} \{e^{-x} - yx\}, \quad y \geqslant 0,$$

т.е.

$$\varphi(y) = \begin{cases} 0, & \text{если} \quad y = 0, \\ -\infty, & \text{если} \quad y > 0. \end{cases}$$

Таким образом, $\sup_{y \in \mathbf{R}_+} \varphi(y) = \varphi(0) = 0$.

Задача D_X имеет решение $y = 0$, в то время как у P_X решение не существует (хотя $\alpha = \beta$).

Пример 5.6. Для $X = \mathbf{R}_+$, $f(x) = -2\sqrt{x}$, $g(x) = x$ находим

$$q(v) = \begin{cases} -2\sqrt{v}, & \text{если} \quad v \geqslant 0, \\ +\infty & \text{в противном случае,} \end{cases} \qquad q(0) = f(0),$$

$$\varphi(y) = \inf_{x \in \mathbf{R}_+} \{-2\sqrt{x} + yx\} = \begin{cases} -\dfrac{1}{y} & \text{при} \quad y > 0, \\ -\infty & \text{при} \quad y = 0, \end{cases}$$

отсюда $\sup_{y \in \mathbf{R}_+} \varphi(y) = 0$.

Задача P_X имеет решение $x^0 = 0$, задача же D_X решения не имеет (хотя $\alpha = \beta = 0$).

Справедлива следующая теорема о существовании решений задачи D_X.

Теорема 5.6. *Пусть значение $q(0)$ функции чувствительности q задачи P_X конечно. Тогда следующие утверждения эквивалентны*:

1) $-y^0 \in \partial q(0)$;

2) $\inf_{x \in G} f(x) = \max_{y \in \mathbf{R}_+^m} \varphi(y) = \varphi(y^0)$.

Если к тому же P_X имеет решение x^0, то первое утверждение эквивалентно

3) $f(x^0) = \min_{x \in G} f(x) = \max_{y \in \mathbf{R}_+^m} \varphi(y) = \varphi(y^0)$.

Доказательство. Для конечного значения $q(0)$ имеем

$$-y^0 \in \partial q(0) \iff q(v) \geqslant q(0) - y^{0\prime} v \quad \forall v \in \mathbf{R}^m,$$

$$\iff \varphi(y^0) = \inf_{v \in \mathbf{R}_+^m} \{q(v) + y^{0\prime} v\} \geqslant q(0).$$

Отсюда и из слабой теоремы двойственности (теорема 5.1) следует эквивалентность первых двух утверждений теоремы. Эквивалентность первого и третьего утверждений в предположении $q(0) = f(x^0)$ очевидна.

5.4. Сильная теорема двойственности

Поскольку

$$\sup_{y \in \mathbf{R}_+^m} L(x, y) = \sup_{y \in \mathbf{R}_+^m} \{f(x) + y'g(x)\} = \begin{cases} f(x), & \text{если} \quad g(x) \leqslant 0, \\ +\infty, & \text{если} \quad g(x) > 0, \end{cases}$$

то второе утверждение теоремы 5.6 можно записать также в виде

$$\inf_{x \in X} \sup_{y \in \mathbf{R}_+^m} L(x, y) = \max_{y \in \mathbf{R}_+^m} \inf_{x \in X} L(x, y).$$

Третье утверждение эквивалентно

$$f(x^0) = \min_{x \in X} \sup_{y \in \mathbf{R}^m_+} L(x, y) = \max_{y \in \mathbf{R}^m_+} \inf_{x \in X} L(x, y) = L(x^0, y^0), \qquad (5.8)$$

что согласно § 4.4 эквивалентно существованию седловой точки $(x^0, y^0)'$ функции Лагранжа L на множестве $X \times \mathbf{R}^m_+$.

Таким образом, глобальный критерий оптимальности из теоремы 4.24 является частным случаем теорем двойственности для задач P_X и D_X.

Теорема 5.7. (*сильная теорема двойственности*). *Пусть значение $q(0)$ функции чувствительности q задачи P_X конечно, функции f, g выпуклы на множестве X и выполнено условие регулярности R_S^X. Тогда*

$$\inf_{x \in G} f(x) = \max_{y \in \mathbf{R}^m_+} \varphi(y) = \varphi(y^0). \qquad (5.9)$$

Если к тому же задача P_X имеет решение x^0, то

$$f(x^0) = \min_{x \in G} f(x) = \max_{y \in \mathbf{R}^m_+} \varphi(y) = \varphi(y^0). \qquad (5.10)$$

Доказательство. По предположению $\alpha = \inf_{x \in G} f(x)$ конечно. Поэтому система

$$f(x) - \alpha < 0, \qquad g(x) \leqslant 0$$

не имеет в X решений. Если выполнено условие R_S^X, то по теореме 2.104 существует $y^0 \in \mathbf{R}^m_+$ такой, что

$$(f(x) - \alpha) + y^0{}'g(x) \geqslant 0 \qquad \forall\, x \in X.$$

Отсюда следует

$$\varphi(y^0) = \inf_{x \in X} \{ f(x) + y^0{}'g(x) \} \geqslant q(0) = \alpha .$$

Учитывая теорему 5.1, получаем (5.9) и (5.10).

Остановимся теперь на вопросе о связи между доказанными выше теоремами.

Теорема 5.6 делает возможной геометрическую интерпретацию сильной теоремы двойственности (а тем самым и глобального критерия оптимальности из теоремы 4.24). Условие, согласно которому задача D_X имеет решение со свойством $\alpha = \beta$, в силу теоремы 5.6 эквивалентно утверждению о том, что множество epi q имеет невертикальную опорную гиперплоскость H в точке $(0, q(0))' \in \mathbf{R}^{m+1}$ (рис. 5.2). Как показывает пример 5.6, это условие отнюдь не обязано выполняться и в предположении выпуклости функции g.

Если же функция f выпукла, то условие регулярности R_S^X обеспечивает существование подобной невертикальной гиперплоскости, опорной к множеству epi q.

Теоремы 4.23, 4.24 и 5.7 были доказаны путем применения конкретной теоремы отделимости в форме утверждения о разрешимости системы выпуклых неравенств. Сформулируем теперь теоремы 5.6 и 5.7 как утверждения об отделимости для некоторых выпуклых множеств.

В силу монотонности функции q для невертикальной гиперплоскости H, опорной к множеству epi q, всегда $y \geqslant 0$, поэтому соотношение $-y^0 \in \partial q(0)$ эквивалентно неравенству $y^0 \geqslant 0$ и "невертикальной разделимости" выпук-

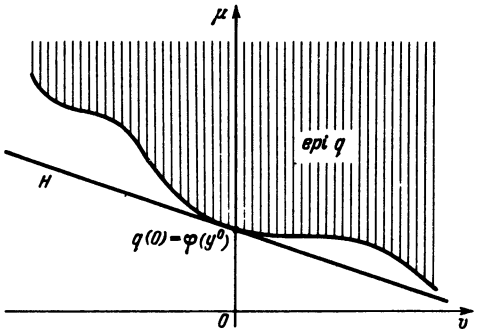


Рис. 5.2.

лых множеств $\operatorname{epi} q - \begin{pmatrix} 0 \\ q(0) \end{pmatrix}$ и $-\mathbf{R}_+^{m+1}$ (рис. 5.3) гиперплоскостью

$$H^0 = \{(v, \mu)' \in \mathbf{R}^{m+1}: \ \mu + y^{0\prime} v = -q^c(y^0)\}.$$

С другой стороны, из доказательства теоремы 5.7 следует разделимость множеств

$$M = \{(v, \mu)' \in \mathbf{R}^{m+1}: \ \exists x \in X: \ f(x) - \alpha \leqslant \mu, \ \ g(x) \leqslant v\} \tag{5.11}$$

и $-\mathbf{R}_+^{m+1}$.

Тесная связь между различными подходами становится ясной из следующего соотношения:

$$\operatorname{int} \operatorname{epi} q \subseteq M + \begin{pmatrix} 0 \\ \alpha \end{pmatrix} \subseteq \operatorname{epi} q. \tag{5.12}$$

Действительно, для $q(v) < +\infty$ неравенство

$$\mu > \inf_{x \in X} \{ f(x): \ \ q(x) \leqslant v \} - \alpha$$

имеет место тогда и только тогда, когда существует вектор $x \in X$ такой, что $f(x) - \alpha \leqslant \mu$ и $g(x) \leqslant v$. Но это означает, что

$$M \supseteq \{(v, \mu)': \ \mu > q(v), \ \ v \in \operatorname{dom} q\} - (0, \alpha)'.$$

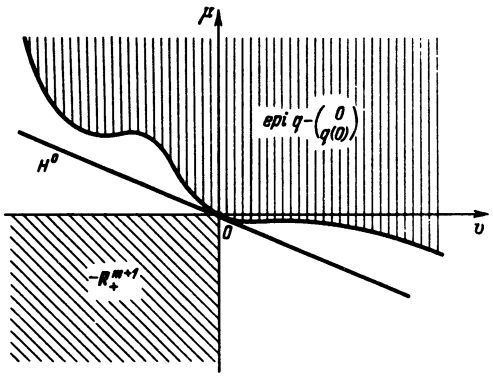


Рис. 5 3.

Так как, с одной стороны,

$$\operatorname{int} \operatorname{epi} q = \{(v, \mu)': \quad \mu > q(v), \quad v \in \operatorname{int} \operatorname{dom} q\},$$

то

$$\operatorname{int} \operatorname{epi} q \subseteq M + \begin{pmatrix} 0 \\ \alpha \end{pmatrix}.$$

С другой стороны, из определения M и q имеем

$$M + \begin{pmatrix} 0 \\ \alpha \end{pmatrix} \subseteq \operatorname{epi} q.$$

Объединяя два последних соотношения, получаем (5.12).

Для задачи выпуклого программирования можно ввести понятие устойчивости. Выпуклая задача P_X называется *устойчивой*, если

$$\partial q(0) \neq \phi \ . \tag{5.13}$$

Устойчивость задачи P_X всегда влечет $G \neq \phi$. Если P_X устойчива, то справедливо одно из двух соотношений:

$$1)\ q(0) = -\infty, \qquad \partial q(0) = \mathbf{R}^m;$$

$$2)\ q(0) = \alpha \text{ конечно и } \phi \neq \partial q(0) \subseteq \mathbf{R}^m. \tag{5.14}$$

Принимая во внимание теоремы 2.78 и 5.6, получаем следующую характеризацию устойчивых задач.

Т е о р е м а 5.8. *Выпуклая задача P_X устойчива тогда и только тогда, когда выполнено одно из следующих условий:*

$$1)\ \exists\, k_0 < 0: \quad q^+(0, h) \geqslant k_0 \quad \forall\, h \in K(\operatorname{dom} q), \quad 0 < \| h \| \leqslant 1;$$

$$2)\ \exists\, k_1 < 0: \quad (q(v) - q(0))/\| v \| \geqslant k_1 \quad \forall\, v \in \operatorname{dom} q, \quad v \neq 0;$$

3) *задача D_X имеет решение y^0 и $\alpha = \beta$.*

Если предполагается существование решения P_X, то устойчивость задачи P_X эквивалентна существованию седловой точки функции Лагранжа L на множестве $X \times \mathbf{R}_+^m$.

Согласно теоремам 5.6 и 5.7 условие регулярности $\mathbf{R}_5^X$ влечет устойчивость задачи P_X. Для частного случая $X = \mathbf{R}^n$ и $I^L = \phi$ этот факт тотчас же следует из теоремы 2.77.

Далее каждое из условий регулярности $R_1 - R_5$ из п. 4.2.3 (при соответствующих предположениях о дифференцируемости) является достаточным для устойчивости задачи выпуклого программирования P_X при $X = \mathbf{R}^n$.

После подробного обсуждения условий, достаточных для выполнения соотношения $\partial q(0) \neq \phi$, мы займемся достаточными условиями для того, чтобы

$$q(0) = q^{cc}(0). \tag{5.15}$$

Разумеется, из $\partial q(0) \neq \phi$ следует справедливость (5.15). Будем искать условия, которые являются более слабыми, нежели существование субдифференциала $\partial q(0)$.

Т е о р е м а 5.9. *Пусть множество $X \subseteq \mathbf{R}^n$ замкнуто и выпукло, функции $f: X \to \mathbf{R}$ и $g: X \to \mathbf{R}^m$ непрерывны и выпуклы, значение $q(0)$ конечно, а множество*

$$M_\gamma = \{x \in G: \quad f(x) \leqslant \gamma\}$$

непусто и компактно для некоторого $\gamma \geqslant q(0)$. *Тогда*

$$\inf_{x \in G} f(x) = \sup_{y \in \mathbf{R}^m_+} \varphi(y).$$

Доказательство. По теореме 3.16 для конечного $q(0)$ соотношение (5.15) выполняется тогда и только тогда, когда выпуклая функция q полунепрерывна снизу в точке $v = 0$. Для того чтобы показать последнее, мы допустим, что q не является полунепрерывной снизу в точке $v = 0$. Тогда найдется последовательность $\{v^k\} \subset \operatorname{dom} q$ такая, что $\|v^k\| \to 0$ и $q(v^k) \to q_0$, где $q_0 < q(0)$. При этом $q(v^k) = \inf_{x \in G_{vk}} f(x)$, где $G_{vk} = \{x \in X: \ g(x) \leqslant v^k\} \neq \phi$. Итак, для каждого $k \in \mathbf{N}$ существует точка $x^k \in G_{vk}$, удовлетворяющая неравенствам

$$q(v^k) \leqslant f(x^k) \leqslant q(v^k) + 1/k.$$

Для определенной таким образом последовательности $\{x^k\}$ имеем $\lim_{k \to \infty} f(x^k) = q_0$.

Выбирая произвольно $\epsilon > 0$, можем для последовательности $\{x^k\}$ определить номер k_ϵ так, чтобы

$$x^k \in M_\epsilon = \{x \in X: \ g_i(x) \leqslant \epsilon, \quad i \in I; \ f(x) \leqslant q(0)\} \quad \forall k \geqslant k_\epsilon. \tag{5.16}$$

По теореме 2.86 из ограниченности множества M_γ следует ограниченность M_ϵ, а тем самым и ограниченность последовательности $\{x^k\}$. Поэтому в $\{x^k\}$ существует подпоследовательность, которая в силу замкнутости множества X и непрерывности функций f и g сходится к точке $x^0 \in X$ со свойством $g(x^0) \leqslant 0$, $f(x^0) = q_0$. Таким образом, применительно к задаче P_X точка x^0 является допустимой и для нее выполняется $f(x^0) < q(0)$. Получили противоречие. Поэтому функция q полунепрерывна снизу в точке $v = 0$. Отсюда с учетом теоремы 5.5 и вытекает требуемое утверждение.

Замечание. Если допустимое множество G задачи P_X непусто и компактно, то и $M_{q(0)+1}$ непусто и компактно. Для многих практических постановок можно считать это предположение выполненным.

Доказанные нами выше теоремы двойственности не дают, конечно, еще полной картины взаимной связи между существованием решений для задач P_X и D_X. В частности, пока остается открытым вопрос о том, как, зная решения двойственной задачи D_X, сделать вывод о решениях прямой задачи P_X. Но прежде чем сформулировать такую "обратную" теорему двойственности, приведем обзор полученных фактов и продемонстрируем возможные случаи на примерах.

5.5. Примеры

В следующих пяти примерах рассматриваются функции двух переменных

$$f(x_1, x_2) = x_2, \qquad g(x_1, x_2) = x_1.$$

Примеры различаются лишь выбором множества $X \subseteq \mathbf{R}^2$.

P_X \ D_X		β конечно				$\beta = \infty$	$\beta = -\infty$
		$\alpha = \beta$		$\alpha > \beta$			
		существует решение	не существует решения	существует решение	не существует решения		
$q(0) = \alpha$ конечно	$\partial q(0) \neq \phi$	пример 5.5	невозможно согласно теореме 5.6				
	$\partial q(0) = \phi$	невозможно согласно теореме 5.6	пример 5.6 пример 5.7	пример 5.10	пример 5.8	невозможно согласно теореме 5.6	пример 5.11
$\alpha = -\infty$		невозможно согласно теореме 5.6					пример 5.3
$\alpha = \infty$				пример 5.2	пример 5.9	пример 5.4	пример 5.1

Пример 5.7. Для (рис. 5.4)

$$X = \{ x \in \mathbf{R}^2 : \ 0 \leqslant x_1 \leqslant 1, \quad 1 - \sqrt{x_1} \leqslant x_2 \leqslant 2 \}$$

получаем (рис. 5.5)

$$\operatorname{dom} q = [0, \infty), \quad q(v) = \begin{cases} 1 - \sqrt{v}, & \text{если } v \in [0, 1), \\ 0, & \text{если } v \in [1, \infty), \end{cases}$$

$$\varphi(y) = \inf_v \{ q(v) + yv \} = \begin{cases} y & \text{если } y \in [0, 1/2), \\ 1 - 1/(4y), & \text{если } y \in [1/2, \infty). \end{cases}$$

Имеем

$$\alpha = q(0) = f(0, 1) = 1 = \beta = \sup_{y \geqslant 0} \varphi(y);$$

задача D_X не имеет решения.

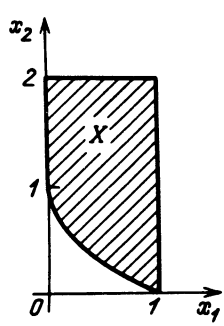


Рис. 5.4.

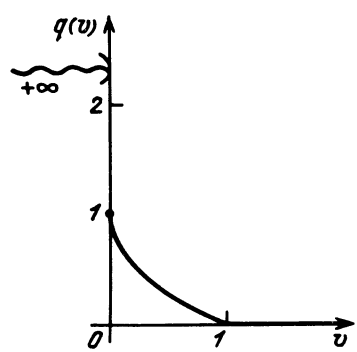


Рис. 5.5.

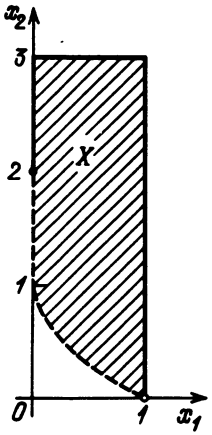


Рис. 5.6.

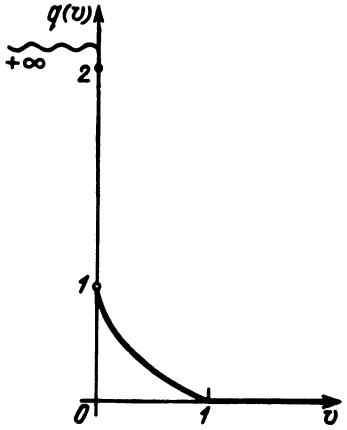


Рис. 5.7.

Пример 5.8. Для (рис. 5.6)

$$X = \{x \in \mathbf{R}^2 : \ 0 \leqslant x_1 \leqslant 1, \quad 2 \leqslant x_2 \leqslant 3 \text{ при } x_1 = 0, \quad 1 - \sqrt{x_1} < x_2 \leqslant 3 \text{ при } x_1 > 0\}$$

получаем (рис. 5.7)

$$\operatorname{dom} q = [0, \infty),$$

$$q(v) = \begin{cases} 2, & \text{если } v = 0, \\ 1 - \sqrt{v}, & \text{если } v \in (0, 1), \\ 0, & \text{если } v \in [1, \infty), \end{cases}$$

$$\varphi(y) = \begin{cases} y, & \text{если } y \in [0, 1/2), \\ 1 - 1/(4y), & \text{если } y \in [1/2, \infty). \end{cases}$$

Имеем

$$\alpha = q(0) = 2 > \beta = 1;$$

задача D_X не имеет решения.

Пример 5.9. Для (рис. 5.8)

$$X = \{x \in \mathbf{R}^2 : \ 0 < x_1 \leqslant 1, \quad 1 - \sqrt{x_1} \leqslant x_2 \leqslant 2\}$$

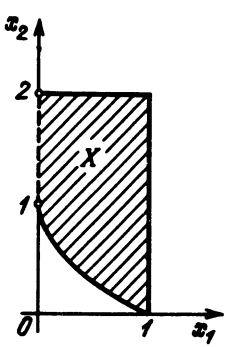


Рис. 5.8.

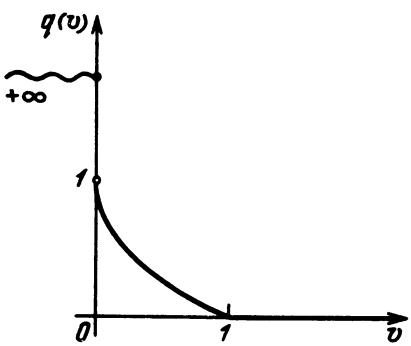


Рис. 5.9.

получаем (рис. 5.9)

$\operatorname{dom} q = (0, \infty)$,

$$q(\upsilon) = \begin{cases} 1 - \sqrt{\upsilon}, & \text{если} \quad \upsilon \in (0, 1), \\ 0, & \text{если} \quad \upsilon \in [1, \infty), \end{cases}$$

$$\varphi(y) = \begin{cases} y, & \text{если} \quad y \in [0, 1/2), \\ 1 - 1/(4y), & \text{если} \quad y \in [1/2, \infty). \end{cases}$$

Имеем

$\alpha = q(0) = +\infty > \beta = 1$;

задача D_X не имеет решения.

Пример 5.10. Для (рис. 5.10)

$X = \{ x \in \mathbf{R}^2 : \ 0 \leqslant x_1 \leqslant 2, \ 3 \leqslant x_2 \leqslant 4 \ \text{при} \ x_1 = 0, \quad 1 < x_2 \leqslant 4 \ \text{при} \ x_1 > 0 \}$

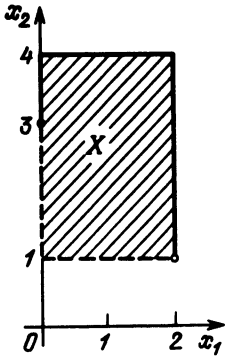


Рис. 5.10.

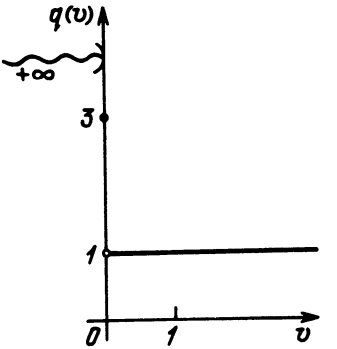


Рис. 5.11.

получаем (рис. 5.11)

$\operatorname{dom} q = [0, \infty)$,

$$q(\upsilon) = \begin{cases} 3, & \text{если} \quad \upsilon = 0, \\ 1, & \text{если} \quad \upsilon \in (0, \infty), \end{cases}$$

$\varphi(y) \equiv 1$ при $y \geqslant 0$.

Имеем

$\alpha = q(0) = 3 > \beta = 1$.

Каждая точка $y \in \mathbf{R}_+$ является решением задачи D_X.

Пример 5.11. Для (рис. 5.12)

$X = \{ x \in \mathbf{R}^2 : \ 0 \leqslant x_1 \leqslant 1, \ 1 \leqslant x_2 \leqslant 2 \ \text{при} \ x_1 = 0, \quad -\infty < x_2 \leqslant 2 \ \text{при} \ x_1 > 0 \}$

получаем (рис. 5.13)

$\operatorname{dom} q = [0, \infty)$,

$$q(\upsilon) = \begin{cases} 1, & \text{если} \quad \upsilon = 0, \\ -\infty, & \text{если} \quad \upsilon \in (0, \infty), \end{cases}$$

$\varphi(y) \equiv -\infty$ при $y \in \mathbf{R}_+$.

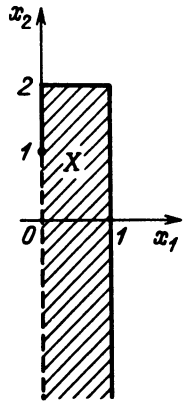


Рис. 5.12.

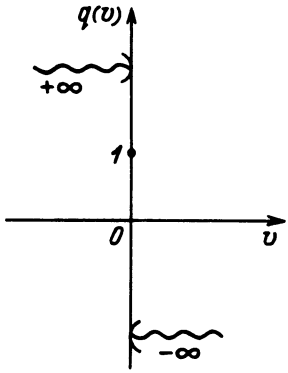


Рис. 5.13.

Имеем

$\alpha = q(0) = 1 > \beta = -\infty$.

Хотя супремум в задаче D_X достигается для любого $y \in \mathbf{R}_+$, несмотря на это мы в этом случае не можем говорить о существовании решений задачи D_X.

5.6. Обратные теоремы двойственности

Для исследования задач P_X и D_X примеры 5.2 и 5.9 интересны в том отношении, что в них задача D_X имеет конечное оптимальное значение, но задача P_X не содержит допустимых точек. Тем не менее рассмотренные в обоих примерах задачи P_X будут совместными при сколь угодно малых "возмущениях". Эту ситуацию точнее характеризует следующая теорема.

Т е о р е м а 5.10. *Пусть заданы задачи P_X и D_X с выпуклой функцией чувствительности q у P_X. Предположим, что $q(0) = +\infty$ и β конечно. Тогда* $0 \in \overline{\operatorname{dom} q}$.

Д о к а з а т е л ь с т в о. Предположим, $0 \notin \overline{\operatorname{dom} q}$. Тогда нулевой вектор может быть строго отделен некоторой гиперплоскостью от выпуклого множества $\operatorname{dom} q$, т.е. существуют $\epsilon > 0$ и $w \in \mathbf{R}^m_+$ такие, что $w'v \geqslant \epsilon > 0$ $\forall v \in \operatorname{dom} q$.

В силу того, что $v = g(x) \in \operatorname{dom} q$ $\forall x \in X$, справедливо, в частности, неравенство $\inf\limits_{x \in X} \{w'g(x)\} > 0$. Если для некоторого $y^* \in \mathbf{R}^m_+$ выполняется соотношение

$$\inf_{x \in X} \{f(x) + y^{*\prime} g(x)\} = \varphi(y^*) > -\infty,$$

то для произвольного $t \geqslant 0$ получим

$$\varphi(y^* + tw) \geqslant \varphi(y^*) + t \inf_{x \in X} \{w'g(x)\}.$$

При $t \to +\infty$ отсюда вытекает $\beta = +\infty$, что противоречит условиям теоремы.

Случай $\beta = -\infty$ (см. также примеры 5.1, 5.3 и 5.11) достаточно полно характеризуется следующей теоремой.

Теорема 5.11. *Пусть заданы задачи P_X и D_X с выпуклой функцией чувствительности q у P_X. Равенство $\beta = -\infty$ эквивалентно тому, что $q(v) = -\infty$ хотя бы для одного $v \in \operatorname{dom} q$.*

Доказательство. 1. Пусть $\varphi(y) = -\infty$ $\forall y \in \mathbf{R}^m_+$ и $q(v) > -\infty$ $\forall y \in \operatorname{dom} q$. По теореме 5.2 $\operatorname{int} \operatorname{dom} q \neq \phi$. Поэтому для произвольно выбранного фиксированного $\bar{v} \in \operatorname{int} \operatorname{dom} q$ существует субградиент $\tilde{y} \in -\mathbf{R}^m_+$. Полагая $\bar{y} = -\tilde{y}$, имеем

$$q(v) \geqslant q(\bar{v}) + \tilde{y}'(v - \bar{v}) \quad \forall v \in \mathbf{R}^m,$$

и тем самым

$$\varphi(\bar{y}) = \inf_v \{ g(v) + \bar{y}'v \} \geqslant q(\bar{v}) + \bar{y}'\bar{v},$$

что противоречит предположению $\varphi(y) = -\infty$ $\forall y \in \mathbf{R}^m_+$.

2. Обратное утверждение следует из теоремы 5.1. Возможен также путь доказательства, опирающийся на теорему 3.16.

Теперь мы хотим, исходя их свойств двойственной задачи D_X, сделать вывод о свойствах прямой задачи P_X. Пусть y^0 — решение задачи D_X: $\varphi(y^0) = \max\limits_{y \in \mathbf{R}^m_+} \varphi(y)$ и

$$Q_0 = \{ x^0 \in \mathbf{R}^n : \varphi(y^0) = L(x^0, y^0) = \inf_{x \in X} \{ f(x) + y^{0\prime} g(x) \} \} . \qquad (5.17)$$

Следует отметить, что множество Q_0 может быть пустым и, кроме того, не каждый элемент из Q_0 является одновременно решением задачи P_X, даже в случае $\alpha = \beta$. Проиллюстрируем это на примере.

Пример 5.12. Для $X = \mathbf{R}$, $g(x) = x$,

$$f(x) = \begin{cases} -x, & \text{если } x \leqslant 0, \\ 0, & \text{если } x \geqslant 0, \end{cases}$$

получаем

$$\inf_{x \in \mathbf{R}} \{ f(x) : g(x) \leqslant 0 \} = f(0) = 0,$$

$$L(x, y) = \begin{cases} -x + yx, & \text{если } x \leqslant 0, \\ yx, & \text{если } x \geqslant 0, \end{cases} \quad y \geqslant 0,$$

$$q(v) = \begin{cases} -v, & \text{если } v \leqslant 0, \\ 0, & \text{если } v \geqslant 0, \end{cases}$$

$$\varphi(y) = \begin{cases} 0, & \text{если } 0 \leqslant y \leqslant 1, \\ -\infty, & \text{если } y > 1, \end{cases}$$

и поэтому

$$\sup_{y \in \mathbf{R}_+} \varphi(y) = \varphi(0) = 0.$$

В силу равенства

$$L(x^0, 0) = \begin{cases} -x^0, & \text{если } x^0 \leqslant 0, \\ 0, & \text{если } x^0 \geqslant 0, \end{cases}$$

имеем $Q_0 = \{ x^0 : x^0 \geqslant 0 \}$, но лишь $x^0 = 0$ является допустимой точкой задачи P_X.

Однако справедлива следующая теорема.

Теорема 5.12. *Пусть* y^0 *– решение задачи* D_X *и*

$$\alpha = \beta = \varphi(y^0). \tag{5.18}$$

Для того чтобы точка $x^0 \in G$ *была решением задачи* P_X*, необходимо и достаточно, чтобы выполнялись соотношения*:

$$x^0 \in Q_0, \quad y^{0\prime} g(x^0) = 0, \quad g(x^0) \leqslant 0.$$

Доказательство. 1. Пусть x^0 – решение задачи P_X. Тогда в силу (5.18) функция Лагранжа L имеет в $(x^0, y^0)'$ седловую точку на множестве $X \times \mathbf{R}^m_+$, а это означает, что

$$\left.\begin{aligned} &L(x^0, y^0) = \min_{x \in X} L(x, y^0), \\ &y^{0\prime} g(x^0) = 0, \quad g(x^0) \leqslant 0. \end{aligned}\right\} \tag{5.19}$$

2. Если выполняются соотношения (5.19), то $(x^0, y^0)'$ – седловая точка функции L на множестве $X \times \mathbf{R}^m_+$ и тем самым x^0 – решение задачи P_X.

Сформулируем "обратную" теорему двойственности.

Теорема 5.13. *Пусть в задаче* P_X *множество* $X \subseteq \mathbf{R}^n$ *замкнуто и выпукло, функции* $f: X \to \mathbf{R}$ *и* $g: X \to \mathbf{R}^m$ *непрерывны и выпуклы на* X*,* y^0 *– решение задачи* D_X*, множество* Q_0 *непусто и компактно и в* Q_0 *существует допустимая в задаче* P_X *точка. Тогда*:

1) *задача* P_X *имеет решение* x^**, для которого*

$$f(x^*) = q(0) = \sup_{y \in \mathbf{R}^m_+} \varphi(y) = \varphi(y^0);$$

2) *если к тому же* $g(x) = \text{const}$ *на* Q_0*, то любая точка* $x^0 \in Q_0$ *является решением* P_X.

Доказательство. 1. Так как по предположению задача D_X имеет решение, а для P_X существует допустимая точка, то достаточно показать, что $\alpha = \beta$ и P_X имеет решение *).

2. Докажем вначале существование решения задачи P_X. Пусть $x^0 \in Q_0$. Функция

$$\psi(x) = L(x, y^0) - \beta, \quad x \in X,$$

выпукла и непрерывна на X, а множество $\{x \in X: \psi(x) \leqslant 0\} = Q_0$ по условию компактно. Поэтому в силу теорем 2.86, 2.87 компактно при $\mu = \alpha - \beta + 1$ и множество

$$Q_\mu = \{x \in X: \psi(x) \leqslant \mu\} = \{x \in X: f(x) + y^{0\prime} g(x) \leqslant q(0) + 1\}.$$

Пусть теперь $\{x^k\}$ – последовательность допустимых в задаче P_X точек и $\lim\limits_{k \to \infty} f(x^k) = q(0)$. Поскольку $x^k \in G \ \forall k \in \mathbf{N}$ и тем самым $y^{0\prime} g(x^k) \leqslant 0$, существует число k_0 такое, что $x^k \in Q_\mu \ \forall k \geqslant k_0$. Инфимум

*) Согласно приведенной на стр. 155 таблице здесь могут встретиться лишь те случаи, которые характеризуются примерами 5.5 и 5.10.

функции f на множествах

$$\{x^k\}_{\substack{k\in\mathbf{N}\\ k\geqslant k_0}} \subseteq Q_\mu \cap G \subseteq G$$

удовлетворяет соотношениям

$$q(0) \geqslant \inf_{x\in Q_\mu\cap G} f(x) \geqslant \inf_{x\in G} f(x) = q(0).$$

Поэтому

$$\alpha = \inf_{x\in G} f(x) = \inf_{x\in Q_\mu\cap G} f(x).$$

Так как множество $Q_\mu \cap G$ компактно, а функция f непрерывна, то инфимум f на G достигается в некоторой точке $x^* \in Q_\mu \cap G \subseteq G$:

$$f(x^*) = \inf_{x\in G} f(x).$$

Тем самым показано существование решения задачи P_X.

3. При проверке соотношения $\alpha = \beta$ мы будем придерживаться (в частности, в обозначениях) доказательства теоремы 5.9. Покажем, что функция q полунепрерывна снизу в точке $0 \in \mathbf{R}^m$. Для доказательства теоремы 5.9 был существенным тот факт, что построенная последовательность $\{x^k\}$ при $k \geqslant k_0$ лежит в компактном множестве, а это обеспечивает существование сходящейся к некоторой допустимой точке x^* подпоследовательности, причем $f(x^*) = q_0$.

Вместо множества M_ϵ, определенного в (5.16), используем теперь множество Q_μ. Очевидно, в силу неравенств $q(v^k) \leqslant f(x^k) \leqslant q(v^k) + 1/k$ и $g(x^k) \leqslant v^k$, где $\|v^k\| \to 0$, найдется номер k_0 такой, что $x^k \in Q_\mu$ при $k \geqslant k_0$. Вследствие компактности множества Q_μ последовательность $\{x^k\}$ ограничена и, как и при доказательстве теоремы 5.9, отсюда мы делаем вывод о существовании допустимой точки x^* со свойством $f(x^*) = q_0 < q(0)$. Получим противоречие.

4. Из равенства $f(x^*) = \varphi(y^0)$ следует, что $(x^*, y^0)'$ является седловой точкой функции L на множестве $X \times \mathbf{R}^m_+$. Отсюда вытекает, что $x^* \in Q_0$ и, если функция g постоянна на Q_0, $f(x^*) = f(x^0)$ $\forall x^0 \in Q_0$. Следовательно, любая точка $x^0 \in Q_0$ является решением P_X.

Теорема 5.13 справедлива, в частности, когда множество Q_0 содержит лишь единственный элемент x^0 (при условии, что выполнены остальные предположения). Точка x^0 является тогда единственным решением задачи P_X.

Такой случай возникает, например, когда функция $f(x) + y^0{}'g(x)$ строго выпукла по x на множестве X или когда для некоторой точки $x^* \in Q_0$ существует окрестность, в которой $f(x) + y^0{}'g(x)$ строго выпукла (тогда x^* – единственный элемент Q_0). Для строгой выпуклости $f(x) + y^0{}'g(x)$ на X или, соответственно, в окрестности точки x^*, как известно, достаточно, чтобы этим свойством обладала функция f или одна из функций g_i, для которой $y_i^0 > 0$.

5.7. Теоремы двойственности для частных задач

5.7.1. Задача линейного программирования. В линейном программировании прямой линейной задаче

$$P_L: \quad \min\{c'x: x \in G_L\}, \tag{5.20}$$

где $G_L = \{x \in \mathbf{R}^n : Ax \geqslant b\}$, $c \in \mathbf{R}^n$, $b \in \mathbf{R}^m$, A – матрица порядка $m \times n$, ставится в соответствие двойственная линейная задача

$$D_L: \quad \max\{b'y: y \in G_L^*\}, \tag{5.21}$$

где $G_L^* = \{y \in \mathbf{R}^m : A'y = c,\ y \geqslant 0\}$.

Наряду с данной в литературе рассматриваются и другие пары двойственных задач линейного программирования. Их допустимые множества приведены в следующей таблице:

	прямая задача	двойственная задача
1	$Ax \geqslant b$	$A'y = c,\ y \geqslant 0$
2	$Ax \geqslant b,\ x \geqslant 0$	$A'y \leqslant c,\ y \geqslant 0$
3	$Ax = b,\ x \geqslant 0$	$A'y \leqslant c$
4	$Ax = b$	$A'y = c$

Теоремы двойственности для пар задач 2, 3 и 4 могут быть получены из теоремы двойственности для пары 1 с помощью простых дополнительных выкладок.

Далее мы покажем, как из развитой нами выше общей теории следуют теоремы двойственности для линейного программирования. Для этого запишем P_X в виде

$$P: \quad \min\{f(x): g(x) \leqslant 0,\ x \in X\},$$

где $X = \mathbf{R}^n$; $g(x) = b - Ax$.

В соответствии с общим подходом найдем двойственную целевую функцию

$$\varphi(y) = \inf_{x \in X}\{f(x) + y'g(x)\} = \inf_{x \in \mathbf{R}^n}\{c'x + y'(b - Ax)\} =$$

$$= \inf_{x \in \mathbf{R}^n}\{y'b + (c' - y'A)x\} =$$

$$= \begin{cases} y'b, & \text{если } A'y = c,\ y \geqslant 0, \\ -\infty & \text{в противном случае.} \end{cases}$$

Итак, в рамках нашего общего подхода задаче P ставится в соответствие двойственная задача

$$D: \quad \max\{\varphi(y): y \in \mathbf{R}_+^m\},$$

которая совпадает с уже известной нам двойственной задачей D_L. Так как в данном случае f и g – выпуклые функции, то множество $\operatorname{epi} q$ выпукло. Для конечного α существует решение линейной задачи P_L,

поскольку, как известно, линейная ограниченная снизу на выпуклом многогранном множестве функция f достигает на нем своего минимума.

Кроме того, для P_L выполнено условие регулярности R_S^X. Поэтому для конечного α найдется также по крайней мере одна точка $y^0 \in \mathbf{R}^m$ такая, что

$$\alpha = f(x^0) = c'x^0 = \varphi(y^0) = b'y^0 = \beta. \tag{5.22}$$

Таким образом, для линейных задач случаи из таблицы на с. 155, характеризуемые примерами 5.8, 5.10 и 5.11, невозможны.

Если записать D как прямую задачу

$$-\min\left\{ y'(-b): \begin{pmatrix} -c \\ c \\ 0 \end{pmatrix} + \begin{pmatrix} A' \\ -A' \\ -E \end{pmatrix} y \leqslant 0 \right\},$$

то по аналогии можно заключить, что для конечного β существуют $y^0 \in \in \mathbf{R}^m$ и $x^0 \in \mathbf{R}^n$ такие, что выполняются равенства (5.22). Поэтому из соображений симметрии из таблицы на стр. 155 следует исключить случаи, характеризуемые примерами 5.6, 5.2 и 5.9.

Таким образом, справедлива следующая теорема.

Т е о р е м а 5.14. *Для пары задач линейного программирования P_L и D_L имеют место следующие утверждения*:

1) *если одна из задач имеет решение, то разрешима и другая*; *при этом* $\alpha = \beta$;

2) *если обе задачи имеют допустимые точки, то они обе разрешимы*; *при этом* $\alpha = \beta$.

Для пары двойственных задач линейного програмирования (в $\mathbf{R}^n$ и соответственно в $\mathbf{R}^m$) не могут возникать разрывы в двойственности.

5.7.2. Задача квадратичного программирования. Задача квадратичного программирования относится к первым нелинейным задачам, для которых были сформулированы теоремы двойственности (например, Дорном [14] в 1960 г.).

Прямой задаче квадратичного программирования

$$P_Q: \min\left\{ f(x) = \frac{1}{2} x'Cx - c'x: x \in G_Q \right\}, \tag{5.23}$$

где $G_Q = \{ x \in \mathbf{R}^n: Ax \leqslant b \}$, $c \in \mathbf{R}^n$, $b \in \mathbf{R}^m$, C – симметрическая положительно полуопределенная матрица порядка $n \times n$, A – матрица порядка $m \times n$, ставится в соответствие двойственная задача квадратичного программирования

$$D_Q: \max\left\{ -\frac{1}{2} x'Cx - b'y: (x, y)' \in G_Q^* \right\}, \tag{5.24}$$

где $G_Q^* = \{ (x, y)' \in \mathbf{R}^n \times \mathbf{R}_+^m: Cx - A'y - c = 0 \}$.

Для того чтобы для этих задач вывести теоремы двойственности из нашей общей теории, положим в задаче P_X

$$X = \mathbf{R}^n, \quad f(x) = \frac{1}{2} x'Cx - c'x, \quad g(x) = Ax - b.$$

Как в линейном случае, справедливо утверждение: если целевая функция f прямой задачи P_Q ограничена снизу на G_Q, то она достигает своего минимума на G_Q в некоторой точке этого множества (см. [61].).

В рамках общего подхода задаче P_Q соответствует двойственная задача

$$\max\{\varphi(y): y \in \mathbf{R}^m_+\} = \max\left\{\inf_{x \in \mathbf{R}^n}\left\{\frac{1}{2}x'Cx - c'x + y'(Ax - b)\right\}: y \in \mathbf{R}^m_+\right\}. \tag{5.25}$$

Для произвольного фиксированного $y \in \mathbf{R}^m_+$ функция Лагранжа

$$L(x, y) = \frac{1}{2}x'Cx - c'x + y'(Ax - b) = \frac{1}{2}x'Cx - y'b + (y'A - c')x,$$

$(x, y)' \in \mathbf{R}^n \times \mathbf{R}^m_+$, ограничена снизу по x тогда и только тогда, когда существует точка $x^0 \in \mathbf{R}^n$ такая, что

$$\nabla_x L(x^0, y) = Cx^0 + A'y - c = 0.$$

Для этой точки

$$L(x^0, y) = -\frac{1}{2}x^{0\prime}Cx^0 - y'b.$$

Поэтому задача (5.25) совпадает с задачей D_Q.

Так же как и в линейном программировании, для квадратичной задачи P_Q (в силу линейности ограничений) всегда выполнено условие регулярности R^X_S. Поэтому, помимо сформулированных ранее общих теорем двойственности, для пары двойственных задач квадратичного программирования P_Q, D_Q справедлива следующая теорема.

Т е о р е м а 5.15. *Пусть заданы задачи P_Q и D_Q. Если значение α конечно, то:*

1) *существует x^0 – решение задачи P_Q;*

2) $\alpha = f(x^0) = \beta$;

3) *существует вектор $y^0 \in \mathbf{R}^m_+$ такой, что $(x^0, y^0)'$ – решение задачи D_Q; при этом $L(x^0, y^0) = \beta$.*

Кроме того, имеет место следующая "обратная" теорема двойственности.

Т е о р е м а 5.16. *Если $(x^0, y^0)'$ – решение двойственной задачи D_Q, то существует точка $x^* \in G_Q$, удовлетворяющая соотношениям:*

1) $C(x^* - x^0) = 0$;

2) $\alpha = \inf\limits_{x \in G_Q} f(x) = f(x^*) = \beta$.

Д о к а з а т е л ь с т в о. Из слабой теоремы двойственности (теорема 5.1) следует, что f ограничена снизу на замкнутом многогранном множестве G_Q. Следовательно, существует x^* – решение задачи P_Q. Тогда по теореме 5.15 имеем $\alpha = \beta$, а по теореме 4.24 $L(x^*, y^0) = \min\limits_x L(x, y^0)$. Отсюда следует, что $x^{*\prime}C + y^{0\prime}A - c' = 0$. Но, кроме того, $x^{0\prime}C + y^{0\prime}A - c' = 0$, а тем самым и $C(x^* - x^0) = 0$.

З а м е ч а н и е. Следует отметить, что условия теоремы 5.16 не гарантируют, что x^0 будет решением задачи P_Q. Однако справедлива следующая теорема.

Теорема 5.17. *Пусть в задаче P_Q матрица C порядка $n \times n$ положительно определена. Если $(x^0, y^0)' \in \mathbf{R}^n \times \mathbf{R}^m_+$ является решением задачи D_Q, то x^0 – решение P_Q и $\alpha = \beta$.*

Доказательство. По условию функция Лагранжа L строго выпукла по x для произвольного фиксированного $y \in \mathbf{R}^m_+$.

Из замечания к теореме 5.13 следует, что все ее условия выполнены. Отсюда получаем требуемое утверждение.

Полагая, в частности, $C = O$ (O – нулевая матрица), из приведенных выше теорем двойственности снова получаем известные теоремы линейного программирования.

Соответствующая модификация предыдущего подхода позволяет вывести теоремы двойственности для пары симметричных двойственных задач квадратичного программирования:

$$P'_Q: \min\left\{ f(x,y) = \frac{1}{2}x'Cx + \frac{1}{2}y'Dy + c'x : (x,y)' \in G'_Q \right\}, \tag{5.26}$$

$$G'_Q = \{ (x,y)' \in \mathbf{R}^n_+ \times \mathbf{R}^m_+ : Ax + Dy \geqslant -b \},$$

$$D'_Q: \max\left\{ L(x,y) = -\frac{1}{2}x'Cx - \frac{1}{2}y'Dy - b'y : (x,y)' \in G^{*\prime}_Q \right\}, \tag{5.27}$$

$$G^{*\prime}_Q = \{ (x,y)' \in \mathbf{R}^n_+ \times \mathbf{R}^m_+ : Cx - A'y \geqslant -c \},$$

где A – матрица порядка $m \times n$, C и D – симметрические положительно определенные матрицы порядка $n \times n$ и $m \times m$ соответственно.

На подробностях мы здесь не останавливаемся.

5.8. Теоремы двойственности для задач с дифференцируемыми выпуклыми функциями ограничений и цели

Пусть в дальнейшем функции $f: \mathbf{R}^n \to \mathbf{R}$ и $g_i: \mathbf{R}^n \to \mathbf{R}$, $i \in I$, выпуклы и дифференцируемы. Рассмотрим двойственные (по Вулфу [85]) задачи

$$P: \min\{ f(x) : x \in G \}, \tag{5.28}$$

$$G = \{ x \in \mathbf{R}^n : g(x) \leqslant 0 \},$$

$$D: \max\{ L(x,y) : (x,y)' \in G^* \}, \tag{5.29}$$

$$G^* = \{ (x,y)' \in \mathbf{R}^n \times \mathbf{R}^m_+ : \nabla_x L(x,y) = 0 \}.$$

Вначале обсудим связь между двойственной задачей D и задачей D_X, которая ставилась в соответствие задаче P в рамках общего подхода:

$$\max\{ \varphi(y) : y \in \mathbf{R}^m_+ \} = \max\{ \inf_{x \in \mathbf{R}^n} L(x,y) : y \in \mathbf{R}^m_+ \}. \tag{5.30}$$

Если в (5.30) инфимум функции $L(x, y^*)$ при фиксированном $y^* \geqslant 0$ достигается в точке x^*, то $\nabla_x L(x^*, y^*) = 0$. Поэтому для любого фиксированного y^* из множества

$$M = \{ y \in \mathbf{R}^m_+ : \inf_{x \in \mathbf{R}^n} L(x,y) \text{ достигается} \} \tag{5.31}$$

соотношение

$$\varphi(y^*) = L(x^*, y^*)$$

выполняется для всех $(x^*, y^*)' \in \{(x, y^*)' \in \mathbf{R}^n \times \mathbf{R}_+^m : \nabla_x L(x, y^*) = 0\}$, так что задача D может быть записана также в форме

$$\max\{\varphi(y) : y \in M\}.$$

Точки $y \notin M$ не оказывают влияния на отыскание максимума в задаче D. Поэтому очевидно, что слабая теорема двойственности (теорема 5.1) выполняется и для задач (5.28) и (5.29).

Следует заметить, что множество M, вообще говоря, не является выпуклым и может быть пусто и при конечном значении β.

Пусть x^0 – решение задачи (5.28). При выполнении условий сильной теоремы двойственности (теорема 5.7) для задач (5.28) и (5.29) существует точка $y^0 \in \mathbf{R}_+^m$ такая, что

$$\alpha = f(x^0) = L(x^0, y^0) = \inf_x L(x, y^0) = \varphi(y^0) = \beta.$$

Очевидно, что $(x^0, y^0)'$ будет тогда решением и задачи (5.29), т.е. мы получили следующий результат (см. [54]).

Теорема 5.18. *Пусть $x^0 \in \mathbf{R}^n$ – решение задачи P и выполнено условие регулярности R_5. Тогда существует точка $y^0 \in \mathbf{R}_+^m$ такая, что $(x^0, y^0)'$ будет решением задачи D и*

$$\alpha = f(x^0) = L(x^0, y^0) = \sup\{L(x, y) : \nabla_x L(x, y) = 0,\ y \in \mathbf{R}_+^m\}.$$

Мангасаряну принадлежит следующая "обратная" теорема двойственности.

Теорема 5.19. *Пусть x^0 – решение задачи P, $(x^*, y^*)'$ – решение задачи D, $L(x, y^*)$ – строго выпуклая по x в некоторой окрестности точки x^* функция и для P выполнено условие регулярности R_5. Тогда*

$$x^* = x^0, \quad \alpha = f(x^0) = L(x^0, y^*) = L(x^*, y^*) = \beta.$$

Доказательство. Из теоремы 5.18 вытекает существование точки $y^0 \in \mathbf{R}_+^m$ такой, что $(x^0, y^0)'$ является решением задачи D, при этом

$$L(x^*, y^*) = L(x^0, y^0) = \sup\{L(x, y) : \nabla_x L(x, y) = 0,\ y \in \mathbf{R}_+^m\}.$$

Кроме того, по теореме 5.12 $y^{0\prime} g(x^0) = 0$.

Если допустить, что $x^* \neq x^0$, то в силу строгой выпуклости $L(x, y^*)$ в окрестности x^* и соотношения $\nabla_x L(x^*, y^*) = 0$ получим

$$L(x^0, y^*) - L(x^*, y^*) > \nabla_x L(x^*, y^*)'(x^0 - x^*) = 0$$

и, далее,

$$L(x^0, y^*) > L(x^*, y^*) = L(x^0, y^0).$$

Отсюда $y^{*\prime} g(x^0) > y^{0\prime} g(x^0) = 0$ и тем самым $y^{*\prime} g(x^0) > 0$. Последнее же противоречит тому, что $y^* \geqslant 0$ и $g(x^0) \leqslant 0$. Таким образом, $x^* = x^0$.

Замечание. Теорема 5.19, очевидно, слабее теоремы 5.13.

Имеются работы, в которых по аналогии с подходом Вулфа [85] и Мангасаряна [58] при специальных предположениях получен ряд других теорем двойственности для задач (5.28) и (5.29). Соответствующий обзор можно найти в статье Эльстера и Зуппе [23].

В качестве примера приведем следующую теорему двойственности, связанную с именами Хэнсона [38] и Хьюарда [41]. Доказательство теоремы проводится без использования общей теории (хотя это и возможно).

Т е о р е м а 5.20. *Пусть задача D разрешима в точке* $(x^0, y^0)'$, *для задачи P выполнено условие регулярности* R_5. *Если имеет место одно из двух условий*:

а) (*условие Хьюарда*) *функции* f, g *дважды непрерывно дифференцируемы,*

матрица $\nabla_x^2 L(x^0, y^0)$ *невырождена*; (5.32)

б) (*условие Хэнсона*) *существуют открытое множество* $Y^0 \subseteq \mathbf{R}^m$, *точка* $y^0 \in Y^0$ *и дифференцируемая функция* p: $Y^0 \to \mathbf{R}^n$ *такие, что*

$$x^0 = p(y^0), \quad \nabla_x L(x, y)\,|_{x = p(y)} = 0 \quad \forall y \in Y^0, \tag{5.33}$$

то x^0 – *решение задачи P и* $f(x^0) = L(x^0, y^0)$.

Д о к а з а т е л ь с т в о. 1. Пусть выполнено условие а). Так как $\nabla_x L(x^0, y^0) = 0$, $|\nabla_x^2 L(x^0, y^0)| \neq 0$, то из теоремы о разрешимости для неявных функций следует существование n-мерной вектор-функции со свойствами, описанными в условии б). Таким образом, достаточно провести доказательство при предположении, что выполнено б).

2. Пусть выполнено условие б). Тогда

$$(x^0, y^0)' = (p(y^0), y^0)' \in \tilde{G} = \{(p(y), y)' :\ y \in Y^0, y \in R_+^m\} \subseteq G^*.$$

В силу $L(x^0, y^0) = \sup\limits_{(x, y)' \in G^*} L(x, y)$ имеем

$$L(p(y^0), y^0) = \sup_{(p(y), y)' \in \tilde{G}} L(p(y), y),$$

т.е. функция $\Psi(y) = L(p(y), y), y \in \tilde{G}$, достигает максимума в точке y^0. Поэтому с необходимостью должно выполняться

$$\nabla_y \Psi(y^0) = \nabla_y p(y^0)\ \nabla_x L(x, y^0)|_{x = p(y^0)} + \nabla_y L(x^0, y^0) \leqslant 0 \tag{5.34}$$

и, как мы покажем ниже,

$$y^{0\prime} \nabla_y \Psi(y^0) = 0. \tag{5.35}$$

В самом деле, если $y^0 = 0$, то справедливость (5.35) очевидна. В случае $y^0 \neq 0$ существует число $\gamma > 0$ такое, что

$$y^0 + \gamma y^0 \in Y^0 \cap \mathbf{R}_+^m, \quad y^0 - \gamma y^0 \in Y^0 \cap \mathbf{R}_+^m.$$

Поэтому из предположения $y^{0\prime} \nabla_y \Psi(y^0) < 0$ с учетом формулы Тейлора следовало бы

$$L(p(y^0 - \gamma y^0), y^0 - \gamma y^0) > L(p(y^0), y^0),$$
$$L(p(y^0 + \gamma y^0), y^0 + \gamma y^0) > L(p(y^0), y^0),$$

а это – противоречие. Таким образом, и для $y^0 \neq 0$ выполняется соотношение (5.35).

Принимая во внимание, что $\nabla_x L(x, y)|_{x = p(y)} = 0$ для $y \in Y^0$, из (5.34) и (5.35) получим

$$g(x^0) \leqslant 0, \quad y^{0\prime} g(x^0) = 0.$$

Так как $(x^0, y^0)' \in G^*$, то справедливо также соотношение

$$\nabla f(x^0) + \nabla g(x^0) y^0 = 0, \quad y^0 \geqslant 0,$$

т.е. для $(x^0, y^0)'$ выполнены условия Куна – Таккера (4.20) и (4.21), а это значит, что x^0 – решение задачи P и

$$f(x^0) = f(x^0) + y^{0\prime} g(x^0) = L(x^0, y^0).$$

З а м е ч а н и е. Численный анализ двойственной задачи (5.29) в целом наталкивается на трудности, вызванные, помимо других причин, тем, что в качестве ограничений выступают нелинейные уравнения, допустимая область G^* не обязательно выпуклая и функция L не всегда вогнута на G^*. Поэтому в современной литературе, в особенности ориентированной на приложения, определенное предпочтение отдается рассмотрению двойственной задачи (5.30).

5.9. Обобщения теоремы двойственности Фенхеля

Развитый в § 5.1 подход к теории двойственности в нелинейном программировании оказался достаточно общим, чтобы вывести критерии оптимальности в форме теорем двойственности для большого числа конкретных задач нелинейного программирования. Тем не менее, существует еще более общая концепция двойственности, основные идеи которой мы изложим ниже, следуя Рокафеллару [74]. Как частный случай при этом получим классическую теорему двойственности Фенхеля.

В дальнейшем будем предполагать, что $C \subseteq \mathbf{R}^n$, $D \subseteq \mathbf{R}^m$ – непустые выпуклые множества, $f\colon C \to \mathbf{R}$ – выпуклая функция, $g\colon D \to \mathbf{R}$ – вогнутая функция, A – матрица порядка $m \times n$.

Прямой задаче выпуклого программирования

$$P_K\colon \min\{f(x) - g(Ax)\colon x \in G_K\}, \tag{5.36}$$

$$G_K = C \cap \{x\colon Ax \in D\} = C \cap A^{-}(D)\ ^{*)},$$

ставится в соответствие *двойственная задача*

$$D_K\colon \max\{g_c(y) - f^c(A'y)\colon y \in G_K^*\}, \tag{5.37}$$

где $G_K^* = D^* \cap A'^{-}(C^*)$, $C^* = D(f^c)$, $D^* = D(g_c)$. Задачи P_K и D_K называются *совместными*, если, соответственно, $G_K \neq \phi$ и $G_K^* \neq \phi$. В противном случае задачи называются *несовместными*.

Аналогично § 5.1 положим

$$\alpha = \inf_{x \in G_K} \{f(x) - g(Ax)\}, \quad \beta = \sup_{y \in G_K^*} \{g_c(y) - f^c(A'y)\} \tag{5.38}$$

---

*) A^{-} обозначает полный прообраз соответствующего множества.

и условимся, что

$$\alpha = +\infty, \quad \text{если } G_K = \phi, \quad \beta = -\infty, \quad \text{если } G_K^* = \phi.$$

Имеет место следующая слабая теорема двойственности.

Т е о р е м а 5.21. *Справедливо неравенство* $\alpha \geqslant \beta$.

Д о к а з а т е л ь с т в о. Утверждение, очевидно, верно, если одна из двух задач несовместна.

Пусть P_K и D_K совместны. Тогда по определению сопряженной функции для $(x, y)' \in G_K \times G_K^*$ выполняются неравенства

$$f(x) + f^c(A'y) \geqslant x'(A'y), \tag{5.39}$$

$$g(Ax) + g_c(y) \leqslant (Ax)'y = x'(A'y), \tag{5.40}$$

откуда следует

$$f(x) + f^c(A'y) \geqslant g(Ax) + g_c(y) \quad \forall (x, y)' \in G_K \times G_K^*,$$

$$f(x) - g(Ax) \geqslant g_c(y) - f^c(A'y) \quad \forall (x, y)' \in G_K \times G_K^*. \tag{5.41}$$

Таким образом, получаем $\alpha \geqslant g_c(y) - f^c(A'y) \quad \forall y \in G_K^*$ и окончательно $\alpha \geqslant \beta$.

Из (5.41) вытекает конечность α и β, если обе задачи P_K и D_K одновременно совместны. Задача D_K соответствующим преобразованием может быть приведена к виду P_K. Имеем

$$\sup_{y \in G_K^*} \{g_c(y) - f^c(A'y)\} = -\inf_{y \in G_K^*} \{f^c(A'y) - g_c(y)\} =$$

$$= -\inf_{y \in G_K^*} \{(-g_c)(y) - (-f^c)(A'y)\}.$$

Задача

$$\tilde{P}_K: \quad \inf_{y \in G_K^*} \{(-g_c)(y) - (-f^c)(A'y)\}$$

будет того же типа, что и P_K. Соответствующая задаче $\tilde{P}_K$ двойственная задача имеет вид

$$\tilde{D}_K: \quad \sup_{x \in \tilde{G}_K^*} \{(-f^c)_c(x) - (-g_c)^c((A')'x)\},$$

где $\tilde{G}_K^* = C^{**} \cap (A')^{'-}(D^{**}) = C^{**} \cap A^{-}(D^{**})$, $C^{**} = D(-f^c)_c$, $D^{**} = D(-g_c)^c$.

Простые выкладки показывают, что

$$(-f^c)_c(x) = -f^{cc}(-x), \quad (-g_c)^c((A')'x) = -g_{cc}(A(-x)),$$

$$C^{**} = -D(f^{cc}), \quad D^{**} = -D(g_{cc}).$$

Поэтому $\tilde{D}_K$ эквивалентна следующей выпуклой задаче:

$$\tilde{P}_K: \quad \inf_{x \in \tilde{G}_K^*} \{f^{cc}(x) - g_{cc}(Ax)\}.$$

Если f, $-g$ — замкнутые снизу выпуклые функции, то

$$f^{cc}(x)=f(x) \quad \forall x \in -C^{**}=C,$$
$$g_{cc}(y)=g(y) \quad \forall y \in -D^{**}=D, \tag{5.42}$$

и, следовательно, $\tilde{P}_K$ эквивалентна задаче P_K.

Отсюда следует возможность формальной дуализации утверждений, относящихся к одной из задач P_K, D_K, если f, $-g$ — замкнутые снизу выпуклые функции.

Оказывается также, что эта возможность в общем случае (если не требовать замкнутости функций) не реализуется, поскольку тогда вместо (5.42) выполняются лишь соотношения

$$f^{cc}(x) \leqslant f(x), \quad g_{cc}(y) \geqslant g(y),$$

а также

$$-C^{**} \supseteq C, \quad -D^{**} \supseteq D.$$

Задачу P_K погрузим в *класс "возмущенных" задач* $P_{K(v)}$ вида

$$P_{K(v)}\colon \min\{f(x)-g(Ax-v)\colon x \in G_{K(v)}\},$$

где $G_{K(v)} = C \cap \{x\colon Ax - v \in D\} = C \cap A^{-}(v+D)$, $v \in \mathbf{R}^m$. При этом $v \in \mathbf{R}^m$ называется *вектором "возмущений"*.

Очевидно,

$$P_{K(0)} = P_K.$$

По аналогии с подходом из § 5.2 вводится *функция чувствительности* $q\colon \mathbf{R}^m \to \bar{\mathbf{R}}$ задачи $P_{K(v)}$:

$$q(v)=\begin{cases} \inf\limits_{x \in G_{K(v)}} \{f(x)-g(Ax-v)\} & \text{при } v \in V, \\ +\infty & \text{при } v \in \mathbf{R}^m \setminus V, \end{cases} \tag{5.43}$$

где

$$V=\{v \in \mathbf{R}^m\colon G_{K(v)} \neq \phi\} = A(C)-D. \tag{5.44}$$

Подобно теореме 5.3 доказывается следующая теорема.

Т е о р е м а 5.22. *Для функции чувствительности* $q\colon \mathbf{R}^m \to \bar{\mathbf{R}}$ *задачи* $P_{K(v)}$ *множество* $\operatorname{epi} q = \{(v, \mu)' \in \mathbf{R}^m \times \mathbf{R}\colon \mu \geqslant q(v)\}$ *выпукло.*

Аналогично определению устойчивости задачи P_X, данному в § 5.4, назовем задачу P_K *устойчивой*, если

$$\partial q(0) \neq \phi.$$

Соответственно задача D_K называется устойчивой, если

$$\partial r(0) \neq \phi,$$

где

$$r(u)=\begin{cases} \inf\limits_{y \in G_K^*} \{f^c(A'y-u)-g_c(y)\} & \text{при } u \in U^*, \\ +\infty & \text{при } u \in \mathbf{R}^n \setminus U^*, \end{cases}$$

$$U^*=\{u \in \mathbf{R}^n\colon G^*_{K(u)} \neq \phi\}.$$

Справедлива следующая сильная теорема двойственности (Рокафеллар [74]).

Т е о р е м а 5.23. *Пусть задана пара двойственных задач* P_K *и* D_K *и* $\alpha < +\infty$. *Имеют место утверждения*:

1) *для устойчивости задачи* P_K *необходимо и достаточно, чтобы* $\alpha = \beta$;

2) *задача* P_K *устойчива тогда и только тогда, когда существует точка* $y^0 \in G_K^*$ *такая, что* $\alpha = \beta = g_c(y^0) - f^c(A'y^0)$;

3) *если* f, $-g$ – *замкнутые выпуклые функции, то для устойчивости задачи* D_K *необходимо и достаточно, чтобы существовала точка* $x^0 \in G_K$ *такая, что* $\alpha = f(x^0) - g(Ax^0) = \beta$.

Д о к а з а т е л ь с т в о. 1. Если $\alpha = -\infty$, то по теореме 5.21 $\alpha = \beta = -\infty$. Задача P_K тогда согласно определению устойчива и $G_K^* = \phi$.

2. Пусть задача P_K устойчива и α конечно. Тогда найдется точка $y^0 \in \partial q(0)$. Покажем, что $\alpha = \beta = g_c(y^0) - f^c(A'y^0)$.

Справедливо соотношение

$$y^0 \in \partial q(0) \iff q(v) \geqslant q(0) + y^{0\prime} v \quad \forall v \in \mathbf{R}^m,$$

а потому

$$\alpha = q(0) = \inf_{v \in \mathbf{R}^m} \{q(v) - y^{0\prime} v\} =$$

$$= \inf_{v \in \mathbf{R}^m} \{ \inf_{x \in G_K(v)} \{f(x) - g(Ax - v)\} - y^{0\prime} v \} =$$

$$= \inf_{v \in \mathbf{R}^m} \{ \inf_{x \in G_{K(v)}} \{(y^{0\prime}(Ax - v) - g(Ax - v)) - ((A'y^0)'x - f(x))\}\} =$$

$$= \inf_{v \in \mathbf{R}^m} \{ \inf_{(x,z)' \in B_v} \{(y^{0\prime} z - g(z)) - ((A'y^0)'x - f(x))\}\} =$$

$$= \inf_{(x,z)' \in \bigcup\limits_{v \in \mathbf{R}^m} B_v} \{(y^{0\prime} z - g(z)) - ((A'y^0)'x - f(x))\} =$$

$$= \inf_{(x,z)' \in C \times D} \{(y^{0\prime} z - g(z)) - ((A'y^0)'x - f(x))\} =$$

$$= \inf_{z \in D} \{y^{0\prime} z - g(z)\} - \sup_{x \in C} \{(A'y^0)'x - f(x)\},$$

где $B_v = \{(x,z)' \in C \times D\colon Ax - z = v\}$.

Так как множества C и D предполагались непустыми, то имеем

$$\gamma = \inf_{z \in D} \{y^{0\prime} z - g(z)\} < +\infty,$$

$$\delta = \sup_{x \in C} \{(A'y^0)'x - f(x)\} > -\infty.$$

Конечность α влечет конечность γ и δ; это значит, что

$$y^0 \in D^*, \quad A'y^0 \in C^*, \quad \gamma = g_c(y^0), \quad \delta = f^c(A'y^0).$$

Таким образом,

$$\alpha = g_c(y^0) - f^c(A'y^0) \leqslant \sup_{y \in G_K^*} \{g_c(y) - f^c(A'y)\} = \beta$$

и, в силу теоремы 5.21,

$$\alpha = g_c(y^0) - f^c(A'y^0) = \beta.$$

3. Если для $y^0 \in G_K^*$ выполняется равенство

$$\alpha = g_c(y^0) - f^c(A'y^0),$$

то необходимо $\alpha = \beta$.

Из приведенной в начале доказательства цепочки равенств, пробегая ее в обратном порядке, получим

$$q(v) \geqslant q(0) + y^{0\prime} v \quad \forall v \in \mathbf{R}^m,$$

т.е. $y^0 \in \partial q(0)$. Поэтому $\partial q(0) \neq \phi$ и, следовательно, задача P_K устойчива.

Тем самым доказаны первое и второе утверждения теоремы.

4. Третье утверждение получим, если примем во внимание эквивалентность задач D_K и $\tilde{P}_K$ и применим уже доказанные утверждения к задаче $\tilde{P}_K$.

Назовем задачу P_K *сильно совместной*, если

$$\mathrm{ri}\, C \cap \{x \in \mathbf{R}^n : Ax \in \mathrm{ri}\, D\} \neq \phi.$$

Сильная совместность является достаточным условием для устойчивости задачи P_K.

Т е о р е м а 5.24. *Если задача P_K сильно совместна, то она устойчива.*

Д о к а з а т е л ь с т в о. В силу сильной совместности задачи P_K имеем

$$0 \in A(\mathrm{ri}\, C) - \mathrm{ri}\, D.$$

С учетом теорем 2.29 и 2.31 получим

$$0 \in A(\mathrm{ri}\, C) - \mathrm{ri}\, D = \mathrm{ri}\, A(C) - \mathrm{ri}\, D = \mathrm{ri}(A(C) - D). \tag{5.45}$$

Теперь либо $q(v) = -\infty \ \forall v \in \mathrm{ri}\,(A(C) - D)$, и потому P_K устойчива, либо функция q конечна на множестве $A(C) - D$, в частности, конечно значение $q(0)$. Применяя теорему 2.49 к множеству $\mathrm{epi}\, q$, получим $\partial q(0) \neq \phi$, т.е. задача P_K устойчива и в этом случае.

Рассмотрим теперь частный случай задач P_K, D_K, полагая в (5.36), (5.37) $m = n$ и $A = E$:

$$P: \ \min\{f(x) - g(x) : x \in C \cap D\}, \tag{5.46}$$

$$D: \ \max\{g_c(y) - f^c(y) : y \in C^* \cap D^*\}. \tag{5.47}$$

Задачи $P_{K(v)}$ теперь имеют вид

$$P_{K(v)}: \ \min\{f(x) - g(x - v) : x \in G_{K(v)}, \ v \in \mathbf{R}^n\},$$

$$G_{K(v)} = C \cap (v + D), \ v \in \mathbf{R}^n.$$

Сильная совместность задачи (5.46) в данном случае просто означает

$\mathrm{ri}\,C \cap \mathrm{ri}\,D \neq \phi$.

Из сильной совместности задачи (5.46) в силу теоремы 5.24 вытекает устойчивость этой задачи. Поэтому из теоремы 5.23 следует классическая теорема двойственности Фенхеля [26].

Т е о р е м а 5.25. *Пусть задачи P и D определены согласно* (5.46) *и* (5.47). *Предположим, что $C, D \subseteq \mathbf{R}^n$ – выпуклые множества,* $\mathrm{ri}\,C \cap \mathrm{ri}\,D \neq \phi$, *$f\colon C \to \mathbf{R}$ – выпуклая, а $g\colon D \to \mathbf{R}$ – вогнутая функция и значение α конечно. Тогда существует точка $y^0 \in C^* \cap D^*$ такая, что*

$$\alpha = \inf_{x \in C \cap D} \{f(x) - g(x)\} = g_c(y^0) - f^c(y^0) = \max_{y \in C^* \cap D^*} \{g_c(y) - f^c(y)\} = \beta. \tag{5.48}$$

Если положить $(f - \alpha)(x) = f(x) - \alpha$, то теорема 5.25 может быть проинтерпретирована геометрически как теорема отделимости для выпуклых множеств $\mathrm{epi}\,(f - \alpha) = \mathrm{epi} f - (0, \alpha)'$ и $\mathrm{hypo}\, g$ в пространстве $\mathbf{R}^n \times \mathbf{R}$. Соотношение (5.48) означает

$(f - \alpha)^c(y^0) = g_c(y^0)$.

Полагая $\gamma = g_c(y^0)$, для всех $x \in C$ и всех $x \in D$ получим

$(f - \alpha)(x) \geqslant y^{0\prime} x - \gamma \geqslant g(x)$.

Таким образом, $y^{0\prime} x - z = \gamma$ – уравнение невертикальной гиперплоскости $H \subseteq \mathbf{R}^{n+1}$, разделяющей множества $\mathrm{epi} f - (0, \alpha)'$ и $\mathrm{hypo}\, g$ (рис. 5.14).

Если, в частности, $\alpha > 0$, то $\inf_{x \in C \cap D} \{f(x) - g(x)\}$ можно интерпретировать как инфимум "вертикального расстояния" между точками $(x, \nu)' \in \mathrm{epi}\, f$ и $(x, \mu)' \in \mathrm{hypo}\, g$, $x \in C \cap D$. Соответственно $\sup_{y \in C^* \cap D^*} \{g_c(y) - f^c(y)\}$ с учетом определения f^c и g_c можно интерпретировать как супремум "вертикального расстояния" между двумя невертикальными гиперплоскостями H_1, H_2, каждая из которых разделяет $\mathrm{epi}\, f$ и $\mathrm{hypo}\, g$ (рис. 5.15). Тем самым теорема 5.25 наглядно выражает следующий очевидный факт: максимальное вертикальное расстояние между двумя параллельными невертикальными гиперплоскостями, которые разделяют множества $\mathrm{epi} f$ и $\mathrm{hypo}\, g$ (в силу условия $\mathrm{ri}\,C \cap \mathrm{ri}\,D \neq \phi$ это расстояние реализуется хотя бы на одной паре гиперплоскостей H_1^*, H_2^*), равно инфимуму вертикального расстояния между точками $(x, \nu)' \in \mathrm{epi} f$ и $(x, \mu)' \in \mathrm{hypo}\, g$, $x \in C \cap D$ (см. рис. 5.15). При этом в каждом из приведенных на рис. 5.14 и 5.15 случаев не существует решения задачи P.

Покажем теперь, как сильная теорема двойственности (теорема 5.7) следует из теоремы 5.23. Считая, что выполнены условия теоремы 5.1, рассмотрим с этой целью задачу

$P_Z\colon \min \{\psi(z)\colon z \in \Gamma\}$,

где $\Gamma = \{z \in Z\colon \chi(z) \leqslant 0\}$, $Z \subseteq \mathbf{R}^s$ – непустое множество, ψ, χ – отображения вида $\psi\colon Z \to \mathbf{R}$, $\chi\colon Z \to \mathbf{R}^t$.

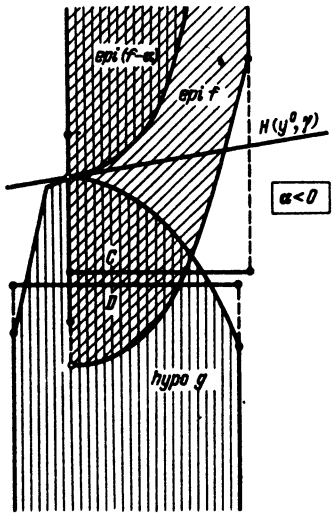


Рис. 5.14.

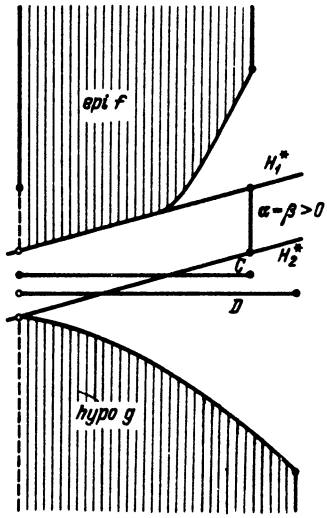


Рис. 5.15.

Задачу P_Z можно переписать в форме P (в соответствии с (5.46)). Для этого положим

$$x=(z,v)'\in\mathbf{R}^s\times\mathbf{R}^t,\quad y=(w,u)'\in\mathbf{R}^s\times\mathbf{R}^t,$$

$$C=\{(z,v)':\ z\in Z,\ \chi(z)+v\leqslant 0\},\quad D=\{(z,v)':\ z\in\mathbf{R}^s,\ v\geqslant 0\},$$

$$f(x)=\psi(z)\quad\forall x\in C;\qquad g(x)=0\quad\forall x\in D.$$

Тогда

$$f^c(y)=\sup_{z\in Z,\ \chi(z)+v\leqslant 0}\{z'w+u'v-\psi(z)\},\quad y\in C^*,$$

$$g_c(y)=0,\quad y\in D^*=\{(w,u)':\ w=0,\ u\geqslant 0\}.$$

Отсюда получаем

$$\sup_{y\in C^*\cap D^*}\{g_c(y)-f^c(y)\}=$$

$$=\sup_{(0,u)'\in C^*,\ u\geqslant 0}\Big\{-\sup_{z\in Z,\ \chi(z)+v\leqslant 0}\{z'w+u'v-\psi(z)\}\Big\}=$$

$$=\sup_{u\geqslant 0}\Big\{\inf_{z\in Z,\ \chi(z)+v\leqslant 0}\{\psi(z)-u'v\}\Big\}=$$

$$=\sup_{u\geqslant 0}\inf_{z\in Z}\{\psi(z)+u'\chi(z)\}.$$

Таким образом, задача, двойственная к P_Z (в смысле (5.47)), соответствует задаче

$$D_Z:\ \max\{\varphi(u):\ u\in\mathbf{R}^t_+\},$$

где $\varphi(u)=\inf_{z\in Z}\{\psi_i(z)+u'\chi(z)\}$.

Если задача P_Z устойчива, то устойчива также эквивалентная P_Z задача (5.46). Пусть значение α для задачи P_Z конечно. Тогда по теореме 5.23 существует вектор $y^0 = (0, u^0)'$ такой, что

$$\alpha = \beta = \inf_{z \in Z} \{ \psi(z) + u^{0\prime} \chi(z) \}.$$

Если для задачи P_Z выполнено условие регулярности R_5 с $I^L \neq \phi$, то соответствующая P_Z задача (5.46) сильно совместна.

5.10. Теоремы о минимаксе и двойственность

Результаты § 5.3, в особенности теорема 5.6, указывают на тесную связь между теорией двойственности и утверждениями о разрешимости задачи о седловой точке. Помимо этого, в § 4.4 обсуждалась эквивалентность задачи о седловой точке и задачи о минимаксе.

В ходе развития теории антагонистических игр появился ряд содержательных теорем о минимаксе. Классическим примером таких теорем является следующая теорема о минимаксе фон Неймана [63].

Т е о р е м а 5.26. *Пусть $X \subseteq \mathbf{R}^n$, $Y \subseteq \mathbf{R}^m$ – компактные множества, функция Φ: $X \times Y \to \mathbf{R}$ при любом фиксированном $y \in Y$ полунепрерывна снизу и выпукла на X, при любом фиксированном $x \in X$ полунепрерывна сверху и вогнута на Y.*

Тогда функция Φ имеет на множестве $X \times Y$ седловую точку $(x^0, y^0)'$, при этом

$$\Phi(x^0, y^0) = \min_{x \in X} \max_{y \in Y} \Phi(x, y) = \max_{y \in Y} \min_{x \in X} \Phi(x, y).$$

Для этой "основной теоремы" теории игр известно несколько доказательств, которые в методическом отношении существенно отличаются друг от друга (см. в этой связи обзор Эрдмана [24]).

Естественно возникает вопрос, насколько широко могут быть использованы известные из теории игр теоремы о минимаксе для получения теорем двойственности в нелинейном программировании. Соответствующие исследования были проведены Гольштейном [34], Карамардианом [43], Стоером [77], Мангасаряном и Понстейном [58], Фогелем [83] и другими. При этом достигнуто определенное обобщение рассмотренной нами выше постановки задачи.

Далее мы приведем доказательство теоремы 5.26, обсудим принципиальные особенности подхода из работы [58] и подробно остановимся на связях с полученными ранее результатами.

Д о к а з а т е л ь с т в о т е о р е м ы 5.26. Для произвольной функции Φ выполняется неравенство (см. доказательство теоремы 4.29)

$$\alpha = \inf_{x \in X} \sup_{y \in Y} \Phi(x, y) \geqslant \sup_{y \in Y} \inf_{x \in X} \Phi(x, y) = \beta. \tag{5.49}$$

Кроме того, из условий теоремы 5.26 следует

$$\alpha = \min_{x \in X} \max_{y \in Y} \Phi(x, y), \quad \beta = \max_{y \in Y} \min_{x \in X} \Phi(x, y).$$

Таким образом, для каждого $x \in X$ справедливо $\max\limits_{y \in Y} \Phi(x, y) \geqslant \alpha$.

Пусть задано $\epsilon > 0$. Для каждого $y \in Y$ определим функцию

$$\psi_y(x) = \Phi(x, y) - \alpha + \epsilon, \quad x \in X.$$

Функция ψ_y полунепрерывна снизу и выпукла на множестве X при любом $y \in Y$. Система выпуклых неравенств

$$\psi_y(x) \leqslant 0, \quad y \in Y,$$

не имеет в X решений. Следовательно, по теореме 2.105 найдутся вектор $p \in \mathbf{R}_+^{n+1}$ и $n+1$ не обязательно попарно различных векторов y^i, при которых функция $\psi(x) = \sum\limits_{i=1}^{n+1} p_i \psi_{y^i}(x)$, $x \in X$, удовлетворяет неравенству $\min\limits_{x \in X} \psi(x) > 0$, т.е. $\min\limits_{x \in X} \{ \sum\limits_{i=1}^{n+1} p_i [\Phi(x, y^i) - \alpha + \epsilon] \} > 0$. Так как функция Φ вогнута по y, то

$$\sum_{i=1}^{n+1} p_i \Phi(x, y^i) \leqslant \Phi\left(x, \sum_{i=1}^{n+1} p_i y^i\right) \quad \forall x \in X.$$

Полагая $\bar{y} = \sum\limits_{i=1}^{n+1} p_i y^i$, имеем $\min\limits_{x \in X} \Phi(x, \bar{y}) > \alpha - \epsilon$ и, далее,

$$\beta = \max_{y \in Y} \min_{x \in X} \Phi(x, y) \geqslant \min_{x \in X} \Phi(x, \bar{y}) > \alpha - \epsilon.$$

Поскольку последнее неравенство справедливо для любого $\epsilon > 0$, то $\beta \geqslant \alpha$. Отсюда с учетом (5.49) получаем $\alpha = \beta$.

Рассмотрим теперь для произвольной функции Φ: $X \times Y \to \mathbf{R}$ и произвольных множеств $X \subseteq \mathbf{R}^n$, $Y \subseteq \mathbf{R}^m$ следующую пару двойственных друг другу задач:

P_V: найти точку $(x^0, y^0)' \in S$, для которой

$$\Phi(x^0, y^0) = \min_{(x, y)' \in S} \Phi(x, y),$$

$$S = \{(x, y)' \in X \times Y: \ \Phi(x, y) = \max_{v \in Y} \Phi(x, v)\}; \tag{5.50}$$

D_V: найти точку $(x^0, y^0)' \in T$, для которой

$$\Phi(x^0, y^0) = \max_{(x, y)' \in T} \Phi(x, y)$$

$$T = \{(x, y)' \in X \times Y: \ \Phi(x, y) = \min_{u \in X} \Phi(u, y)\}. \tag{5.51}$$

Задачи P_V и D_V являются не только несколько более общими, нежели рассмотренные выше задачи, но и имеют то преимущество, что их формулировки более симметричны.

Далее мы рассмотрим задачу о седловой точке:

S_V: найти точку $(x^0, y^0)' \in X \times Y$ такую, что

$$\Phi(x^0, y) \leqslant \Phi(x^0, y^0) \leqslant \Phi(x, y^0) \quad \forall (x, y)' \in X \times Y. \tag{5.52}$$

В случаях $S = \phi$ и $T = \phi$ мы условимся, что $\inf\limits_{(x, y)' \in S} \Phi(x, y) = +\infty$ и $\sup\limits_{(x, y)' \in T} \Phi(x, y) = -\infty$ соответственно. Но даже если $S \neq \phi$, $T \neq \phi$, то, как легко догадаться, задачи P_V, D_V не обязательно имеют решения.

С учетом этой договоренности как аналог слабой теоремы двойственности 5.1 (или теоремы 5.21) из (5.49) вытекает следующая теорема.

Т е о р е м а 5.27. *Для произвольной функции* Φ: $X \times Y \to \mathbf{R}$ *справедливо неравенство*

$$\inf_{(x, y)' \in S} \Phi(x, y) \geqslant \sup_{(x, y)' \in T} \Phi(x, y).$$

Сформулируем теперь теорему о соотношениях между двойственными задачами P_V, D_V и задачей о седловой точке S_V.

Т е о р е м а 5.28. *Для произвольной функции* Φ: $X \times Y \to \mathbf{R}$ *имеют место следующие утверждения*:

1) *любое решение задачи о седловой точке* S_V *является одновременно решением задач* P_V *и* D_V;

2) *если* $(x^1, y^1)'$ *– решение задачи* P_V, $(x^2, y^2)'$ *– решение задачи* D_V *и* $\Phi(x^1, y^1) = \Phi(x^2, y^2)$, *то* $(x^1, y^2)'$ *будет решением задачи о седловой точке* S_V.

В частности, любое общее решение задач P_V *и* D_V *будет в то же время и решением задачи* S_V.

Д о к а з а т е л ь с т в о. 1. Пусть $(x^0, y^0)'$– решение задачи S_V. Тогда $(x^0, y^0)' \in S \times T \neq \phi$. Для произвольной точки $(x, y)' \in S$ имеем

$$\Phi(x, y) \geqslant \Phi(x, y^0) \geqslant \Phi(x^0, y^0),$$

т.е. $(x^0, y^0)'$ – решение задачи P_V.

Соответственно для произвольной точки $(x, y)' \in T$ имеем

$$\Phi(x, y) \leqslant \Phi(x^0, y) \leqslant \Phi(x^0, y^0),$$

т.е. $(x^0, y^0)'$ – решение и задачи D_V.

2. Из соотношений

$$\Phi(x^1, y^1) = \max_{y \in Y} \Phi(x^1, y) \geqslant \Phi(x^1, y^2),$$

$$\Phi(x^2, y^2) = \min_{x \in X} \Phi(x, y^2) \leqslant \Phi(x^1, y^2)$$

в силу равенства $\Phi(x^1, y^1) = \Phi(x^2, y^2)$ следует

$$\max_{y \in Y} \Phi(x^1, y) = \Phi(x^1, y^2) = \min_{x \in X} \Phi(x, y^2).$$

Для вывода дальнейших утверждений необходимо сузить класс рассматриваемых задач. Это достигается путем наложения соответствующих дополнительных ограничений на функцию Φ и на множества X и Y.

По аналогии с предыдущим подходом потребуется выполнение некоторых условий регулярности. Таковыми здесь будут служить введенные Мангасаряном и Понстейном [58] свойства "верхней оценки" (HVP) и "нижней оценки" (LVP) *).

Пусть функция Φ: $X \times Y \to \mathbf{R}$ непрерывна **) и выпукло-вогнута на $X \times Y$, где $X \subseteq \mathbf{R}^n$, $Y \subseteq \mathbf{R}^m$ – замкнутые выпуклые множества. Тогда требуемые условия формулируются следующим образом.

(HVP): Функция Φ удовлетворяет в точке $(\tilde{x}, \tilde{y})' \in X \times Y$ условию (HVP), если существуют число $\rho_1 > 0$ и компактное множество $\tilde{X} \subseteq X$, $\tilde{x} \in \tilde{X}$, такие, что неравенство

$$\Phi(\tilde{x}, \tilde{y}) \geqslant \max_{y \in S(\tilde{y}, \rho_1) \cap Y} \{ \min_{x \in \tilde{X}} \Phi(x, y) \}$$

выполняется тогда и только тогда, когда

$$\Phi(\tilde{x}, \tilde{y}) \geqslant \min_{x \in \tilde{X}} \Phi(x, y) \quad \forall y \in S(\tilde{y}, \rho_1) \cap Y.$$

(LVP): Функция Φ удовлетворяет в точке $(\tilde{x}, \tilde{y})' \in X \times Y$ условию (LVP), если существуют число $\rho_2 > 0$ и компактное множество $\tilde{Y} \subseteq Y$, $\tilde{y} \in \tilde{Y}$, такие, что неравенство

$$\Phi(\tilde{x}, \tilde{y}) \leqslant \min_{x \in S(\tilde{x}, \rho_2) \cap X} \{ \max_{y \in \tilde{Y}} \Phi(x, y) \}$$

выполняется тогда и только тогда, когда

$$\Phi(\tilde{x}, \tilde{y}) \leqslant \max_{y \in \tilde{Y}} \Phi(x, y) \quad \forall x \in S(\tilde{x}, \rho_2) \cap X.$$

Все входящие в условия (HVP) и (LVP) максимумы и минимумы достигаются. Функция Φ на каждом из множеств $\tilde{X} \times (S(\tilde{y}, \rho_1) \cap Y)$ и $\tilde{Y} \times (S(\tilde{x}, \rho_2) \cap X)$ удовлетворяет условиям теоремы о минимаксе фон Неймана (теоремы 5.26). Поэтому

$$\Phi(\tilde{x}, \tilde{y}) \geqslant \min_{x \in X} \max_{y \in S(\tilde{y}, \rho_1) \cap Y} \Phi(x, y) \tag{5.53}$$

и одновременно

$$\Phi(\tilde{x}, \tilde{y}) \leqslant \max_{y \in Y} \min_{x \in S(\tilde{x}, \rho_2) \cap X} \Phi(x, y). \tag{5.54}$$

---

*) Говоря об условиях (HVP) и (LVP), надо отметить, что речь идет о модификации введенного Стоером [77] "B-свойства".

**) Непрерывность понимается здесь (как и далее в этом параграфе) относительно индуцированной топологией пространства $\mathbf{R}^n \times \mathbf{R}^m$ топологии в аффинной оболочке множества $X \times Y$.

Значение введенных условий станет ясным из следующих двух теорем о седловой точке.

Теорема 5.29. *Пусть непрерывная функция* Φ: $X \times Y \to \mathbf{R}$ *выпукло-вогнута на* $X \times Y$, *где* $X \subseteq \mathbf{R}^n$, $Y \subseteq \mathbf{R}^m$ – *замкнутые выпуклые множества. Предположим, что для* $(\tilde{x}, \tilde{y})' \in X \times Y$ *выполняется равенство*

$$\Phi(\tilde{x}, \tilde{y}) = \min_{x \in X} \Phi(x, \tilde{y}). \tag{5.55}$$

Для того чтобы существовал вектор $x^0 \in X$ *такой, что* $(x^0, \tilde{y})'$ *будет седловой точкой функции* Φ *на множестве* $X \times Y$, *необходимо и достаточно, чтобы* Φ *удовлетворяла в точке* $(\tilde{x}, \tilde{y})'$ *условию* (*HVP*). *Кроме того,*

$$\Phi(x^0, \tilde{y}) = \Phi(\tilde{x}, \tilde{y}). \tag{5.56}$$

Доказательство. 1. Равенство $\Phi(x^0, \tilde{y}) = \Phi(\tilde{x}, \tilde{y})$ сразу же следует из свойств седловой точки и (5.55).

2. Пусть $(x^0, \tilde{y})'$ – седловая точка функции Φ на множестве $X \times Y$. Если выбрать $\tilde{X} = \{x^0\}$ и произвольное число $\rho_1 > 0$, то неравенство

$$\Phi(\tilde{x}, \tilde{y}) = \Phi(x^0, \tilde{y})' \geqslant \min_{x \in \tilde{X}} \Phi(x, y) \qquad \forall y \in S(\tilde{y}, \rho_1) \cap Y$$

будет выполняться тогда и только тогда, когда функция Φ удовлетворяет в точке $(\tilde{x}, \tilde{y})'$ условию (*HVP*).

3. Пусть функция Φ удовлетворяет в $(\tilde{x}, \tilde{y})'$ условию (*HVP*). Тогда по теореме 5.26 существует точка $(x^0, y^0)' \in \tilde{X} \times (S(\tilde{y}, \rho_1) \cap Y)$, которая является седловой точкой функции Φ на множестве $\tilde{X} \times (S(\tilde{y}, \rho_1) \cap Y)$, т.е.

$$\Phi(x^0, y) \leqslant \Phi(x^0, y^0) \leqslant \Phi(x, y^0) \qquad \forall x \in \tilde{X}, \qquad \forall y \in S(\tilde{y}, \rho_1) \cap Y, \tag{5.57}$$

$$\Phi(x^0, y^0) = \max_{y \in S(\tilde{y}, \rho_1) \cap Y} \min_{x \in X} \Phi(x, y). \tag{5.58}$$

Из условия (*HVP*) и соотношения (5.57) получаем $\Phi(\tilde{x}, \tilde{y}) \geqslant \Phi(x^0, y^0) \geqslant \Phi(x^0, \tilde{y})$. В силу (5.55) справедливо $\Phi(x^0, \tilde{y}) \geqslant \Phi(\tilde{x}, \tilde{y})$. Отсюда следует $\Phi(\tilde{x}, \tilde{y}) = \Phi(x^0, \tilde{y}) = \Phi(x^0, y^0)$. Применим снова (5.55). Получим

$$\Phi(x^0, \tilde{y}) \leqslant \Phi(x, \tilde{y}) \qquad \forall x \in X. \tag{5.59}$$

В силу (5.57) выполняется неравенство

$$\Phi(x^0, y) \leqslant \Phi(x^0, \tilde{y}) \qquad \forall y \in S(\tilde{y}, \rho_1) \cap Y. \tag{5.60}$$

Остается показать, что соотношение (5.60) имеет место для всех $y \in Y$. Для этого выберем произвольно $y \in Y \setminus (S(\tilde{y}, \rho_1) \cap Y)$. Тогда существует точка $y^1 \in S(\tilde{y}, \rho_1) \cap Y$ и число $\mu \in (0, 1)$ такие, что $y^1 = \mu y + (1 - \mu)\tilde{y}$. Так как функция Φ вогнута по y, то соотношение

$$\Phi(x^0, \tilde{y}) \geqslant \Phi(x^0, y^1) \geqslant \mu\Phi(x^0, y) + (1 - \mu)\Phi(x^0, \tilde{y})$$

выполняется тогда и только тогда, когда $\Phi(x^0, \tilde{y}) \geqslant \Phi(x^0, y)$, т.е. когда (5.60) выполняется для всех $y \in Y$.

Из (5.59) и (5.60) следует, что $(x^0, y^0)'$ – седловая точка функции Φ на множестве $X \times Y$.

Аналогично теореме 5.29 может быть доказана следующая теорема.

Т е о р е м а 5.30. *Пусть непрерывная функция* Φ: $X \times Y \to \mathbf{R}$ *вогнуто-выпукла на* $X \times Y$, *где* $X \subseteq \mathbf{R}^n$, $Y \subseteq \mathbf{R}^m$ – *замкнутые выпуклые множества. Предположим, что для* $(\tilde{x}, \tilde{y})' \in X \times Y$ *выполняется равенство*

$$\Phi(\tilde{x}, \tilde{y}) = \max_{y \in Y} \Phi(\tilde{x}, y). \tag{5.61}$$

Для того чтобы существовал вектор $y^0 \in Y$ *такой, что* $(\tilde{x}, y^0)'$ *будет седловой точкой функции* Φ *на множестве* $X \times Y$, *необходимо и достаточно, чтобы* Φ *удовлетворяла в точке* $(\tilde{x}, \tilde{y})'$ *условию* (LVP). *Кроме того,*

$$\Phi(\tilde{x}, y^0) = \Phi(\tilde{x}, \tilde{y}). \tag{5.62}$$

З а м е ч а н и е. Следует особо подчеркнуть, что при выполнении условий теоремы 5.29 и 5.30 нельзя гарантировать, что $x^0 = \tilde{x}$ и соответственно $y^0 = \tilde{y}$.

После этих предварительных результатов мы можем сформулировать теоремы двойственности для задач P_V и D_V.

Т е о р е м а 5.31. *Пусть непрерывная функция* Φ: $X \times Y \to \mathbf{R}$ *выпукло-вогнута на* $X \times Y$, *где* $X \subseteq \mathbf{R}^n$, $Y \subseteq \mathbf{R}^m$ – *замкнутые выпуклые множества. Справедливы следующие утверждения.*

1) *Если* $(x^0, y^0)'$ – *решение задачи* P_V, *то вектор* $\tilde{y} \in Y$ *такой, что* $(x^0, \tilde{y})'$ *будет одновременно решением задач* P_V, D_V *и* S_V, *существует тогда и только тогда, когда функция* Φ *удовлетворяет в точке* $(x^0, y^0)'$ *условию* (LVP). *Кроме того, имеет место равенство* $\Phi(x^0, y^0) = \Phi(x^0, \tilde{y})$.

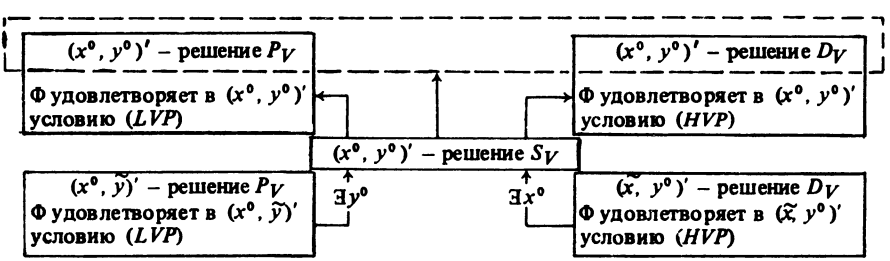
2) *Если* $(x^0, y^0)'$ – *решение задачи* D_V, *то вектор* $\tilde{x} \in X$ *такой, что* $(\tilde{x}; y^0)'$ *будет одновременно решением задач* P_V, D_V *и* S_V, *существует тогда и только тогда, когда функция* Φ *удовлетворяет в точке* $(x^0, y^0)'$ *условию* (HVP). *Кроме того,*

$$\Phi(x^0, y^0) = \Phi(\tilde{x}, y^0).$$

Д о к а з а т е л ь с т в о. 1. Пусть $(x^0, y^0)'$ – решение задачи P_V. Тогда $\Phi(x^0, y^0) = \sup_{y \in Y} \Phi(x^0, y)$. Из теоремы 5.30 следует, что вектор $\tilde{y} \in Y$ такой, что $(x^0, \tilde{y})'$ будет решением задачи S_V (и при этом $\Phi(x^0, \tilde{y}) = \Phi(x^0, y^0)$), существует тогда и только тогда, когда функция Φ удовлетворяет в точке $(x^0, y^0)'$ условию (LVP). Согласно теореме 5.28 $(x^0, \tilde{y})'$ будет тогда решением задач P_V и D_V.

2. Доказательство второго утверждения теоремы проводится аналогично.

Резюмируя предыдущие результаты, получим следующую общую схему (см. Мангасарян и Понстейн [58]):



Для последующих теорем двойственности потребуются более жесткие предположения.

Теорема 5.32. *Пусть непрерывная функция* Φ: $X \times Y \to \mathbf{R}$ *выпукло-вогнута на* $X \times Y$, *где* $X \subseteq \mathbf{R}^n$, $Y \subseteq \mathbf{R}^m$ – *замкнутые выпуклые множества.*

1) *Если* $(x^0, y^0)'$ – *решение задачи* P_V *и функция* $\Phi(x^0, y)$ *строго вогнута в некоторой окрестности* $U(y^0)$ *точки* y^0, *то* $(x^0, y^0)'$ *будет решением задач* D_V *и* S_V.

2) *Если* $(x^0, y^0)'$ – *решение задачи* D_V *и функция* $\Phi(x, y^0)$ *строго выпукла в некоторой окрестности* $U(x^0)$ *точки* x^0, *то* $(x^0, y^0)'$ *будет решением задач* P_V *и* S_V.

Доказательство. Из соображений симметрии можно ограничиться доказательством первого утверждения, которое проведем в три этапа. Вначале покажем, что множество

$$\tilde{Y} = \{ y \in Y:\ \Phi(x^0, y) \geqslant \Phi(x^0, y^0) - \tau;\ \tau > 0,$$

τ – произвольное фиксированное $\}$

компактно и выпукло. Далее проверим, что функция Φ удовлетворяет в точке $(x^0, y^0)'$ условию $(L\,VP)$. И наконец, утверждение теоремы будет следовать из теоремы 5.31.

1. В силу непрерывности функции Φ множество $\tilde{Y}$ замкнуто. По теореме 2.86 $\tilde{Y}$ ограничено. Поэтому $\tilde{Y}$ компактно. Выпуклость $\tilde{Y}$ следует из теоремы 2.55.

2. Пусть число $\rho_2 > 0$ таково, что

$$|\Phi(x, y) - \Phi(x^0, y)| \leqslant \tau/2 \qquad \forall x \in S(x^0, \rho_2) \cap X \qquad \forall y \in \tilde{Y}. \tag{5.63}$$

Так как по условию функция Φ непрерывна, то ρ_2 существуст.

Из непрерывности Φ следует, далее, что для любого $\bar{y} \in Y \setminus \tilde{Y}$ найдется число $\lambda \in (0, 1)$, при котором

$$\Phi(x^0, \lambda y^0 + (1 - \lambda)\bar{y}) = \Phi(x^0, y^0) - \tau, \tag{5.64}$$

т.е.

$$\lambda y^0 + (1 - \lambda)\bar{y} \in \tilde{Y}. \tag{5.65}$$

Из вогнутости функции Φ по y имеем

$$\Phi(x, \lambda y^0 + (1 - \lambda)\bar{y}) \geqslant \lambda\Phi(x, y^0) + (1 - \lambda)\Phi(x, \bar{y}) \qquad \forall x \in X.$$

Отсюда с учетом (5.63), (5.64) и (5.65) следует

$$\Phi(x, y^0) \geqslant \Phi(x^0, y^0) - \tau/2 = \Phi(x^0, \lambda y^0 + (1 - \lambda)\bar{y}) + \tau/2 \geqslant$$
$$\geqslant \Phi(x, \lambda y^0 + (1 - \lambda)\bar{y}) \geqslant \lambda\Phi(x, y^0) + (1 - \lambda)\Phi(x, \bar{y}),$$

т.е.

$$\Phi(x, y^0) \geqslant \Phi(x, \bar{y}) \qquad \forall x \in S(x^0, \rho_2) \cap X \qquad \forall \bar{y} \in Y \setminus \tilde{Y},$$

$$\max_{y \in Y} \Phi(x, y) = \max_{y \in \tilde{Y}} \Phi(x, y) \qquad \forall x \in S(x^0, \rho_2) \cap X.$$

Поскольку множество Y компактно, оба максимума в последнем равенстве существуют.

По предположению $(x^0, y^0)'$ – решение задачи P_V. Поэтому

$$\Phi(x^0, y^0) \leqslant \sup_{y \in Y} \Phi(x, y) = \max_{y \in Y} \Phi(x, y) = \max_{y \in \widetilde{Y}} \Phi(x, y)$$

$\forall x \in S(x^0, \rho_2) \cap X$.

Таким образом, функция Φ удовлетворяет в точке $(x^0, y^0)'$ условию (LVP).

3. Первое утверждение теоремы 5.31 обеспечивает существование точки $y^* \in Y$ такой, что $(x^0, y^*)'$ будет одновременно решением задач P_V, D_V и S_V, а это влечет $\Phi(x^0, y^0) = \Phi(x^0, y^*)$.

В силу строгой вогнутости $\Phi(x^0, y)$ в некоторой окрестности $U(y^0)$ точки y^0 должно выполняться равенство $y^0 = y^*$. Этим теорема доказана полностью.

Из полученных выше результатов (теоремы 5.27 – 5.32) можно вывести ряд теорем о седловой точке и теорем двойственности для конкретных классов задач нелинейного программирования.

В качестве примера покажем, как из P_V и D_V следуют рассмотренные в § 5.8 задачи (5.28) и (5.29) и связанные с ними утверждения (см. Вулф [85]).

Положим $X = \mathbf{R}^n$, $Y = \mathbf{R}^m_+$,

$$\Phi(x, y) = L(x, y) = f(x) + y'g(x), \tag{5.66}$$

где f: $\mathbf{R}^n \to \mathbf{R}$, g: $\mathbf{R}^n \to \mathbf{R}^m$ – дифференцируемые выпуклые функции. Тогда задача P_V сводится к задаче

P: найти вектор $x^0 \in \mathbf{R}^n$ такой, что

$$f(x^0) = \min_{x \in G} f(x), \qquad G = \{x \in \mathbf{R}^n : g(x) \leqslant 0\}.$$

Из общей задачи D_V получаем задачу

D: найти точку $(x^0, y^0)' \in T$ такую, что

$$L(x^0, y^0) = \max_{(x, y)' \in T} L(x, y),$$

$$T = \{(x, y)' \in \mathbf{R}^n \times \mathbf{R}^m_+ : \nabla_x L(x, y) = 0\}.$$

Задачи P и D как раз и есть пара двойственных задач (5.28) и (5.29).

Из задачи о седловой точке S_V получим обычную задачу о седловой точке

S: найти точку $(x^0, y^0)' \in \mathbf{R}^n \times \mathbf{R}^m_+$ такую, что

$$f(x^0) + y'g(x^0) \leqslant f(x^0) + y^{0\prime}g(x^0) \leqslant f(x) + y^{0\prime}g(x) \qquad \forall (x, y)' \in \mathbf{R}^n \times \mathbf{R}^m_+.$$

Если $(x^0, y^0)'$ – решение задачи S, то, очевидно,

$$g(x^0) \leqslant 0, \tag{5.67}$$

т.е. x^0 – допустимая точка прямой задачи P.

Далее, равенство

$$f(x^0) + y^{0\prime}g(x^0) = \inf_{x \in \mathbf{R}^n} \{f(x) + y^{0\prime}g(x)\}$$

эквивалентно соотношению

$$\nabla f(x^0) + \nabla g(x^0) y^0 = 0, \tag{5.68}$$

т.е. $(x^0, y^0)'$ – допустимая точка двойственной задачи D.

Очевидно также, что

$$y^{0\prime} g(x^0) = 0. \tag{5.69}$$

Теорема 5.27 переходит в слабую теорему двойственности 5.21. Теорема 5.28 содержит в соответствии с результатами из § 5.3, 5.4 следующие утверждения:

1) если $(x^0, y^0)' \in \mathbf{R}^n \times \mathbf{R}^m_+$ – решение задачи S о седловой точке, то x^0 будет решением P, а $(x^0, y^0)'$ – решением D;

2) если $x^0 \in \mathbf{R}^n$ – решение задачи P, $(x^0, y^0)' \in \mathbf{R}^n \times \mathbf{R}^m_+$ – решение D, то $(x^0, y^0)'$ будет решением задачи S о седловой точке.

Произвольное решение $(x^0, y^0)'$ задачи D, вообще говоря, не обладает тем свойством, что x^0 будет решением P, и вследствие этого может не являться одновременно решением задачи о седловой точке S.

Теорема 5.29 при наложенных выше ограничениях переходит в следующее утверждение.

Т е о р е м а 5.29′. *Пусть функция* L: $\mathbf{R}^n \times \mathbf{R}^m_+ \to \mathbf{R}$ *задана согласно* (5.66), *точка* $(\tilde{x}, \tilde{y})' \in \mathbf{R}^n \times \mathbf{R}^m_+$ *удовлетворяет соотношению*

$$f(\tilde{x}) + \tilde{y}' g(\tilde{x}) = \inf_{x \in \mathbf{R}^n} \{f(x) + \tilde{y}' g(x)\}. \tag{5.70}$$

Для того чтобы существовал вектор $x^0 \in \mathbf{R}^n$ *такой, что* $(x^0, \tilde{y})'$ *будет седловой точкой функции* L *на множестве* $\mathbf{R}^n \times \mathbf{R}^m_+$, *необходимо и достаточно, чтобы* L *удовлетворяла в точке* $(\tilde{x}, \tilde{y})'$ *условию* (*HVP*). *Кроме того,*

$$L(x^0, \tilde{y}) = L(\tilde{x}, \tilde{y}).$$

Соотношение (5.70) означает, что $(\tilde{x}, \tilde{y})'$ является точкой, для которой выполняется правое неравенство из определения седловой точки.

Принимая во внимание утверждения

1) если $(x^0, y^0)'$ – решение задачи S, то x^0 – решение задачи P;

2) если $f(\tilde{x}) + \tilde{y}' g(\tilde{x}) = \sup_{y \in \mathbf{R}^m_+} \{f(\tilde{x}) + y' g(\tilde{x})\}$, то $g(\tilde{x}) \leqslant 0$ и $\tilde{y}' g(\tilde{x}) = 0$,

теорему 5.30 можно сформулировать в следующем виде.

Т е о р е м а 5.30′. *Пусть* $\tilde{x}$ *– решение задачи* P, *а точка* $\tilde{y} \in \mathbf{R}^m_+$ *такова, что* $\tilde{y}' g(\tilde{x}) = 0$. *Для существования вектора* $y^0 \in \mathbf{R}^m_+$ *такого, что* $(\tilde{x}, y^0)'$ *будет седловой точкой функции* L *на множестве* $\mathbf{R}^n \times \mathbf{R}^m_+$, *необходимо и достаточно, чтобы* L *удовлетворяла в точке* $(\tilde{x}, \tilde{y})'$ *условию* (*LVP*).

В силу предположения о выпуклости функций f и g в задаче P множество $\operatorname{epi} q$ выпукло.

Из сравнения с теоремой 5.7 видно, что для того чтобы задача P имела решение $\tilde{x}$ и была устойчивой, необходимо и достаточно, чтобы P имела решение $\tilde{x}$ и существовал вектор $\tilde{y} \in \mathbf{R}^m_+$ такой, что $\tilde{y}' g(\tilde{x}) = 0$ и функция L удовлетворяет в $(\tilde{x}, \tilde{y})'$ условию (*LVP*).

Кроме того, из замечаний к теореме 5.8 вытекает, что если задача P имеет решение $\tilde{x}$ и выполнено условие регулярности R_1, то существует вектор

$\tilde{y} \in \mathbf{R}_+^m$, $\tilde{y}'g(\tilde{x}) = 0$, такой, что функция L удовлетворяет в $(\tilde{x}, \tilde{y})'$ условию (LVP).

Аналогично получаем следующие два утверждения.

Т е о р е м а 5.31′. *Если* $(x^0, y^0)' \in \mathbf{R}^n \times \mathbf{R}_+^m$ *– решение задачи D, то вектор* $\tilde{x} \in \mathbf{R}^n$ *такой, что* $(\tilde{x}, y^0)'$ *будет одновременно решением задач D и S, а* $\tilde{x}$ *– решением задачи P, существует тогда и только тогда, когда функция L удовлетворяет в точке* $(x^0, y^0)'$ *условию* (HVP). *Кроме того,*

$$L(x^0, y^0) = L(\tilde{x}, y^0) = f(x^0).$$

Т е о р е м а 5.32′. *Если* $(x^0, y^0)'$ *– решение задачи D, а функция* $L(x, y^0)$ *строго выпукла в некоторой окрестности* $U(x^0)$ *точки* x^0, *то* x^0 *– решение задачи P, а* $(x^0, y^0)'$ *– решение задачи S.*

Если задача P имеет решение $\tilde{x}$ и выполнено одно из условий регулярности $R_1 - R_5$, то согласно вышесказанному существует вектор $\tilde{y} \in \mathbf{R}_+^m$, для которого справедливо равенство $\tilde{y}'g(\tilde{x}) = 0$, и функция L удовлетворяет в точке $(\tilde{x}, \tilde{y})'$ условию (LVP). Тогда по теореме 5.30′ найдется вектор $y^0 \in \mathbf{R}_+^m$ такой, что $(\tilde{x}, y^0)'$ будет решением задачи S о седловой точке. Последнее как раз составляет содержание теоремы 5.19.

Из теоремы 5.32′ следует, очевидно, теорема 5.19. Теорема 5.32 показывает также, что предположения теоремы 5.19 о существовании $x^0 \in \mathbf{R}^n$ – решения задачи P – и выполнении условия регулярности R_5 являются излишними, поскольку они следуют из остальных условий.

Наконец, следует заметить, что при выполнении условия Хьюарда (см. (5.32)) утверждение теоремы 5.20 тоже непосредственно вытекает из теоремы 5.32′.

6.1. Основные понятия

В дальнейшем мы будем рассматривать задачу

P: $\min\{f(x)\colon x \in G\}$,

где $G = \{x \in \mathbf{R}^n\colon x \in X,\ g_i(x) \leqslant 0,\ i \in I\}$, $f\colon X \to \mathbf{R}$, $g_i\colon X \to \mathbf{R}$, $i \in I$; $X \subseteq \left(\bigcap_{i \in I} D(g_i)\right) \cap D(f)$.

В § 4.3 были выведены критерии, при помощи которых можно судить о том, является ли точка $x^0 \in X$ элементом множества G^0 решений задачи P или нет. Эти критерии можно использовать также и для отыскания решения x^0 задачи P, решая, например, соответствующим методом систему, определяющую условия Куна – Таккера (4.20), (4.21).

Для конкретных классов задач нелинейного программирования (в частности, для задач квадратичного программирования) такой подход привел к появлению практически реализуемых методов решения ([49, 61]). В общем же случае целесообразно пытаться решать непосредственно задачу P. Как мы увидим далее, при обосновании соответствующих методов большое значение снова будут иметь конкретные критерии оптимальности для задачи P.

Метод решения задачи можно представить как итеративный процесс, в котором, исходя из начальной точки $x^1 \in G_1$, получают последовательность $\{x^k\}$ согласно правилу

$$q(x^k) = x^{k+1} \in X \qquad \forall k \geqslant 1, \tag{6.1}$$

где $G \subseteq G_1 \subseteq X$, а отображение $q\colon X \to X$ и множество G_1 определяются конкретным методом. Упорядоченную пару (q, G_1) назовем *алгоритмом*. Как покажет следующий пример, в некоторых алгоритмах отображение q, помимо точки x^k, зависит также от номера k или от предшествующих точек $x^1, \ldots, x^{k-1}$.

Пример 6.1. Рассмотрим задачу

$\min\{f(x);\ x \in G\}$,

где $G = [a_0, b_0] \subseteq X \subseteq D(f)$, множество X открыто и функция $f\colon X \to \mathbf{R}$ дифференцируема. Пусть $G_1 = G$, $x^1 = (a_0 + b_0)/2$ и, далее,

$$a_k = \begin{cases} x^k & \text{при} \quad f'(x^k) < 0, \\ a_{k-1} & \text{при} \quad f'(x^k) \geqslant 0, \end{cases} \qquad b_k = \begin{cases} b_{k-1} & \text{при} \quad f'(x^k) < 0, \\ x^k & \text{при} \quad f'(x^k) \geqslant 0, \end{cases}$$

$$x^{k+1} = (a_k + b_k)/2 \qquad \forall k \geqslant 1.$$

Несколько шагов алгоритма отмечено на рис. 6.1.

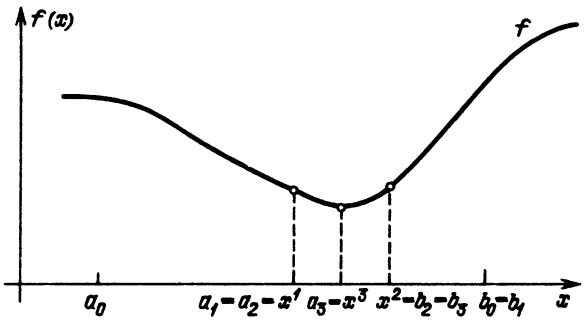


Рис. 6.1.

Желательно, чтобы элемент x^{k_0} последовательности точек $\{x^k\}$, порождаемой алгоритмом (q, G_1), принадлежал множеству решений задачи P. Однако для многих алгоритмов и задач нелинейного программирования этому требованию удовлетворить нельзя. Поэтому наложим более слабое ограничение, а именно потребуем, чтобы точки x^k с достаточно большим номером k сколь угодно точно аппроксимировали решение $x^0 \in G^0$ задачи P относительно метрики, определенной в $\mathbf{R}^n$. В переносном смысле мы говорим тогда о *сходимости алгоритма*: алгоритм (q, G_1) для задачи P называется сходящимся, если порождаемая им последовательность $\{x^k\}$ удовлетворяет следующим условиям:

$K1$: если $x^k \notin G^0 \quad \forall k \geqslant 1$ и $\{x^{k_\nu}\}$ – сходящаяся подпоследовательность из $\{x^k\}$, то $\lim\limits_{\nu \to \infty} x^{k_\nu} \in G^0$;

$K2$: если существует номер k_0, для которого $x^{k_0} \in G^0$, то $x^{k+1} \in G^0$ $\forall k \geqslant k_0$.

Почти все известные алгоритмы решения задачи P основываются на "локальном" сравнении значений целевой функции в некоторой окрестности итерационной точки x^k $(k \geqslant 1)$. Поэтому без дополнительных предположений о задаче P порождаемая алгоритмом (q, G_1) последовательность $\{x^k\}$ может сходиться только к точке локального минимума. Следовательно, условия $K1$ и $K2$ лишь тогда заведомо будут выполнены, когда любая точка локального минимума одновременно является (глобальным) решением задачи P. Таким свойством обладают, например, задачи выпуклого программирования.

Алгоритм (q, G_1), который при некоторых предположениях о функции f удовлетворяет условиям $K1$ и $K2$, называется *методом решения* (для) задачи P.

Если выполняется условие $K2$, то часто желательным является представление метода решения как конечного алгоритма. С этой целью мы дополним (6.1) следующим образом:

$$q(x^k) = x^{k+1} \in X \qquad \forall k \geqslant 1, \qquad \forall x^k \notin G^0. \tag{6.1'}$$

Итак, для метода решения либо выполняется условие $K1$, либо порождается конечная последовательность $\{x^1, x^2, \ldots, x^{k_0}\}$ и x^{k_0} – решение задачи P.

Второй случай возникает, например, при применении лексикографического симплекс-метода к задачам линейного программирования.

Для выполнения условия $K1$ требуется существование у $\{x^k\}$ сходящейся подпоследовательности. При доказательстве сходимости конкретного метода это требование чаще всего удовлетворяют выбором компактного множества $X' \subseteq X$ такого, что $\{x^k\} \subseteq X'$.

Если метод решения задачи P обладает свойством

$$f(x^{k+1}) \leqslant f(x^k), \quad x^k \in X \quad \forall k \geqslant 1,$$

то мы назовем его *методом спуска.*

Многие методы нелинейного программирования являются методами спуска. Однако, как показывает следующий пример, даже для задачи выпуклого программирования P и монотонно убывающей последовательности $\{f(x^{k_\nu})\}$ существование сходящейся последовательности не является достаточным для того, чтобы $\lim\limits_{\nu \to \infty} x^{k_\nu} \in G^0$. Разумеется, в этом случае нельзя говорить о методе решения.

Пример 6.2. Для задачи

$$\min\{f(x) = -x: \ x \in G\}, \quad G = [0, 4] \subset \mathbf{R},$$

положим

$$x^1 = 0, \quad x^{k+1} = q(x^k) = x^k + 1/2^{k-1} \quad \forall k \geqslant 1.$$

Будем иметь

$$\lim_{k \to \infty} x^k = x^1 + \lim_{k \to \infty} \sum_{l=0}^{k-2} \left(\frac{1}{2}\right)^l = 2,$$

$$f(x^k) > f(x^{k+1}) > -2 > \min_{x \in [0,4]} (-x) = -4.$$

За последние годы были предприняты усиленные попытки разработать единую теорию сходимости методов решения задач нелинейного программирования ([69, 86]). Основой для этого явилось понимание алгоритма как точечно-множественного отображения и исследование его свойств непрерывности (см. § 6.8).

Предыдущие рассуждения о сходимости методов решения не связаны с вопросами вычислительной устойчивости. Последняя зависит от поведения погрешности исходных данных, которая неизбежно возникает, например, при округлении координат точек и требуемых значений функций, что обусловлено конечной фиксированной длиной слова вычислительной машины. Например, рекуррентное правило $x^{k+1} = q(x^k)$ может так усилить исходную погрешность, которая вначале казалась безобидной, что она погасит теоретически существующую сходимость (см. в этой связи замечания к доказательству теоремы 6.6). К настоящему времени об устойчивости методов решения имеются почти исключительно эмпирические результаты ([69, 40]), которые с трудом поддаются обобщению. О теоретических исследованиях этого вопроса для некоторых методов можно прочесть в работе [97].

Если мы предположим, что метод устойчив в вычислительном отношении, то требуемое для решения задачи время счета зависит, помимо искус-

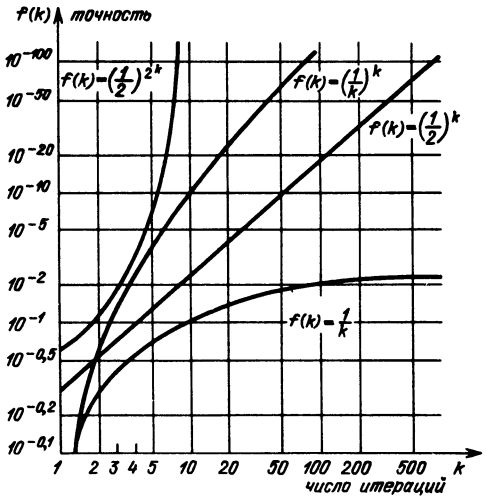


Рис. 6.2.

ства программиста, также от скорости, с которой метод сходится к решению. С этой целью определяется понятие порядка сходимости последовательности $\{x^k\}$.

Будем говорить, что последовательность $\{x^k\} \subseteq \mathbf{R}^n$ сходится к x^0 с *порядком сходимости* p, если p есть наибольшее неотрицательное число, для которого существует $\beta < \infty$ такое, что

$$\overline{\lim_{k \to \infty}} \frac{\| x^{k+1} - x^0 \|}{\| x^k - x^0 \|^p} = \beta.$$

Число β называется *множителем сходимости*.

В случае $p = 1$ и $\beta \in (0, 1)$ последовательность $\{x^k\}$ называется *линейно сходящейся к* x^0 с множителем сходимости β.

При $p = 1$ и $\beta = 0$ последовательность $\{x^k\}$ называется *сверхлинейно сходящейся* к x^0 и при $p = 2$ – *квадратично сходящейся* к x^0.

Пример 6.3. а) Последовательность $\{x^k\} = \{\xi^k\}$ при $0 < \xi < 1$ линейно сходится к $x^0 = 0$ с множителем сходимости $\beta = \xi$. б) Последовательность $\{x^k\} = \{\xi^{2^k}\}$ при $0 < \xi < 1$ квадратично сходится к $x^0 = 0$. в) Последовательность $\{x^k\} = \{1/k\}$ имеет порядок сходимости $p = 1$, но в силу $\beta = 1$ не является линейно сходящейся к x^0. г) Последовательность $\{x^k\} = \{(1/k)^k\}$ сверхлинейно сходится к x^0. Поведение сходимости этих последовательностей наглядно показано на рис. 6.2 (в случаях а) и в) выбрано $\xi = 1/2$).

Если для последовательности $\{x^k\}$ найдутся действительные числа $\gamma \in (0, 1)$ и $C \in [0, \infty)$, для которых выполняется оценка

$$\| x^k - x^0 \| \leqslant C\gamma^k \qquad \forall k \geqslant 1,$$

то будем говорить, что $\{x^k\}$ сходится к x^0 со скоростью геометрической прогрессии со знаменателем γ. В то время как порядок сходимости является свойством, внутренне присущим методам решения, константы, входящие в оценки скорости сходимости (например, C и γ), зависят от задачи.

Кроме оценок скорости сходимости последовательностей $\{x^k\}$, представляют также интерес оценки скорости сходимости последовательности соответствующих значений функции $\{f(x^k)\}$ к оптимальному значению $f(x^0)$.

Для теорем о скорости сходимости чаще всего требуются более сильные предположения о задаче P, чем просто для доказательства сходимости метода.

6.2. Методы решения задач без ограничений

Рассмотрим вначала методы решения задач оптимизации вида

$$\min \{f(x): \ x \in X\}, \tag{6.2}$$

где $X \subseteq D(f) \subseteq \mathbf{R}^n$, f: $D(f) \to \mathbf{R}$, которые мы назовем *задачами без ограничений.* Отдельное обсуждение таких методов целесообразно по разным причинам. Прежде всего, методы решения задач без ограничений являются составной частью методов решения общих задач математического программирования. Так, например, для большого числа методов нелинейного программирования требуется минимизация (или максимизация) функции на одномерном множестве. Если решение x^0 задачи P из § 6.1 удовлетворяет условию $x^0 \in \operatorname{int} G$, то, очевидно, может быть применен (при некоторых предположениях о задаче P и начальной точке x^1) метод решения для задач без ограничений. Далее, некоторые приемы, используемые при доказательстве сходимости и оценках скорости сходимости, применимы и для исследования методов решения задач с ограничениями. Наконец, опыт, приобретаемый при программировании и тестовых испытаниях методов без ограничений, весьма полезен при численной реализации методов для задач с ограничениями.

В последние 15 лет теория методов решения задач без ограничений получила широкое развитие. Подробное изложение этой теории можно найти в работах [40, 56, 66, 69, 97, 101].

В последующем обзоре принято деление методов решения задач без ограничений на отдельные классы по наивысшему порядку участвующих производных функции f.

Методы прямого поиска. Здесь речь пойдет о методах решения задачи (6.2), в которых отображение q для нахождения $x^{k+1} = q(x^k)$, кроме x^k, зависит лишь от определенного числа предшествующих точек $x^l (1 \leqslant l < k)$, а также от соответствующих значений функции $f(x^k)$ и $f(x^l)$ $(1 \leqslant l < k)$.

Преимуществом методов прямого поиска является то, что они применимы к большому классу функций, от которых не требуется таких свойств, как непрерывность и дифференцируемость. Недостаток же их — более медленная скорость сходимости по сравнению с методами, которые работают с первыми и вторыми производными функции f.

Практическое значение методы прямого поиска имеют прежде всего для задач с малым числом переменных, в частности, для минимизации функций одной переменной.

Пусть заданы f: $X \to \mathbf{R}$, $X = [a, b] \subseteq \mathbf{R}$, и мы ищем решение задачи (6.2). Очевидно, выбирая единственную точку $\xi^1 \in X$, а также определяя $f(\xi^1)$, мы не получим никакой информации о собственном подмножестве X, содержащем x^0. Напротив, определение двух значений $f(\xi^1)$ и $f(\xi^2)$, где $\xi^1, \ \xi^2 \in (a, b)$, при условии, что f достигает лишь единственного локального минимума, дает возможность выделить из (a, b) подынтервал, в котором должно лежать решение x^0. Возникает вопрос, как выбирать k точек из (a, b), чтобы подынтервал, в котором лежит x^0, был как можно меньше. Как ответ на этот вопрос, в § 6.4 будут рассмотрены методы Фибоначчи и метод золотого сечения для отыскания минимума функции f: $[a, b] \to \mathbf{R}$.

В качестве примера метода прямого поиска для функций многих переменных приведем одну из модификаций метода покоординатного спуска.

Пример 6.4. Рассмотрим следующий метод.

Шаг 1. Выбрать $x^1 \in X$, $r > 0$, $\epsilon > 0$.

Шаг 2. Положить $k = 1$, $r = r_1$, $m = 1$.

Шаг 3. Определить минимальное $l \in \{m, \ldots, n\}$ таким образом, чтобы для единичного координатного вектора e^l выполнялось

$$\min\{f(x^k + r_k e^l),\ f(x^k - r_k e^l)\} < f(x^k). \tag{6.3}$$

Если такое l не существует, то определить (в случае $m > 1$) минимальное $l \in \{1, \ldots, m-1\}$, при котором (6.3) выполняется.

Если (6.3) не выполняется ни при каких $l \in \{1, \ldots, n\}$, то перейти к шагу 4. В противном случае перейти к шагу 5.

Шаг 4. Положить $r_k = r_k/2$. Если $r_k < \epsilon$, то перейти к шагу 6, в противном случае – к шагу 3.

Шаг 5. Положить $x^{k+1} = x^k + \delta r_k e^l$, $r_{k+1} = r_k$,

$$\delta = \begin{cases} 1 & \text{при } f(x^k + r_k e^l) < f(x^k - r_k + r_k e^l), \\ -1 & \text{при } f(x^k + r_k e^l) \geqslant f(x^k - r_k e^l), \end{cases}$$

$k = k + 1$, $m = l$; перейти к шагу 3.

Шаг 6. Положить $x^0 = x^k$ и остановиться.

В качестве критерия окончания счета применено условие на нижнюю границу для длины шага: $r_k > \epsilon > 0$.

Работу данного метода поясним на задаче

$$\min\{f(x) = (x_1 - 2)^2 + (x_2 - 2)^2 : x \in \mathbf{R}^2\}.$$

Выберем в качестве начальной точки $x^1 = (-0{,}75, 0{,}75)'$, в качестве начальной длины шага $r = 1$, в качестве оценки точности счета $\epsilon = 0{,}2$. Получающиеся итерационные точки и соответствующие значения целевой функции содержатся в табл. 6.1 (см. также рис. 6.3).

Несмотря на то, что методы прямого поиска не используют производные, недифференцируемость функции f может препятствовать сходимости метода (в смысле данного в § 6.1 определения). В этой связи рассмотрим следующий пример.

Пример 6.5. Рассмотрим задачу $\min\{f(x) = \max\{|x_1|, |x_2|\} : x \in \mathbf{R}^2\}$. Функция f, называемая чебышевской нормой, очевидно, достигает своего минимума в точке $x^0 = 0$ с $f(x^0) = 0$ (рис. 6.4). Описанный в примере 6.4 метод будет применим, пока x^k не принадлежит множеству $M = \{x \in \mathbf{R}^2 \setminus \{0\} : x_1^2 = x_2^2\}$. В случае $x^k \in M$ метод не находит ни одной точки $x \in \mathbf{R}^2$, для которой $f(x) < f(x^k)$, и поэтому останавливается, не достигнув x^0, каким бы малым ни было выбрано $\epsilon > 0$. Такая ситуация не может встретиться, если функция f дифференцируема на множестве M и все ее частные производные не равны нулю одновременно.

Таблица 6.1

k	x_1	x_2	$f(x)$	k	x_1	x_2	$f(x)$
1	– 0,75	0,75	9,125	5	2,25	1,75	0,125
2	0,25	0,75	4,625	6	2,25	2,00	0,0625
3	1,25	0,75	2,125	7	2,00	2,00	0
4	2,25	0,75	1,625				

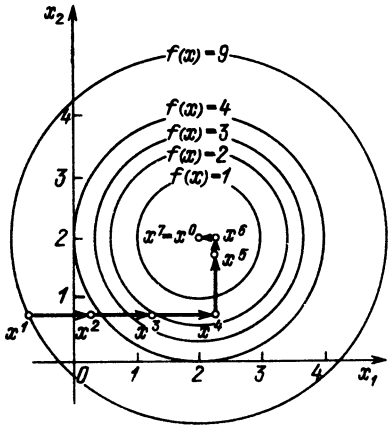


Рис. 6.3.

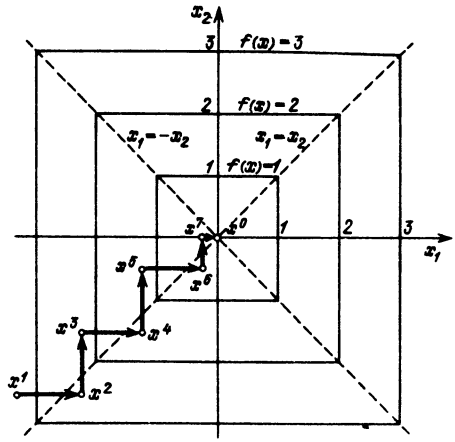


Рис. 6.4.

Другие методы прямого поиска можно найти в работах [2, 10, 40, 56, 97].

Градиентные методы. Пусть в задаче (6.2) $X = \mathbf{R}^n$ и функция f: $\mathbf{R}^n \to \mathbf{R}$ непрерывно дифференцируема. Предположим, что алгоритм $(q, \mathbf{R}^n)$ имеет структуру

$$x^{k+1} = x^k + r_k z^k, \qquad k \geqslant 1, \tag{6.4}$$

где вектор z^k называется *направлением движения* из точки x^k, а число r_k – *длиной шага.*

Если z^k удовлетворяет условию $\nabla f(x^k)' z^k < 0$, то для достаточно малого r_k соотношение (6.4) задает метод спуска, поскольку в силу разложения

$$f(x^{k+1}) = f(x^k) + r_k \nabla f(x^k)' z^k + o(\| z^k r_k \|)$$

для достаточно малого r_k очевидным образом имеем $f(x^{k+1}) < f(x^k)$.

Если, в частности, положим $z^k = -\nabla f(x^k)$, то получим рекуррентное правило

$$x^{k+1} = x^k - r_k \nabla f(x^k), \quad k \geqslant 1, \tag{6.5}$$

являющееся базовым для градиентных методов.

В *методе Коши* (методе наискорейшего спуска) длина шага выбирается согласно правилу

$$f(x^k - r_k \nabla f(x^k)) = \min_{r \geqslant 0} f(x^k - r \nabla f(x^k)). \tag{6.6}$$

Сходимость метода Коши описывается следующим утверждением ([98, 101]).

Т е о р е м а 6.1. 1) *Пусть функция* f: $\mathbf{R}^n \to \mathbf{R}$ *непрерывно дифференцируема и ограничена снизу*; *градиент* ∇f *удовлетворяет условию Липшица с константой* K_1,

$$\| \nabla f(x) - \nabla f(y) \| \leqslant K_1 \| x - y \| \quad \forall x, y \in \mathbf{R}^n. \tag{6.7}$$

Тогда $\| \nabla f(x^k) \| \to 0$ *при* $k \to \infty$ *независимо от выбора начальной точки* x^1.

2) *Пусть функция* $f: \mathbf{R}^n \to \mathbf{R}$ *дважды непрерывно дифференцируема и*

$$m \| y \|^2 \leqslant y' \nabla^2 f(x) y \leqslant M \| y \|^2 \qquad \forall x, y \in \mathbf{R}^n \tag{6.8}$$

при некоторых $M \geqslant m > 0$.

Тогда последовательность $\{x^k\}$ *при любом выборе начальной точки* x^1 *сходится к решению* x^0 *и для скорости сходимости справедливы оценки*

$$\| x^{k+1} - x^0 \| \leqslant C \left(\frac{M-m}{M+m} \right)^{k+1}, \qquad C = \left(\frac{M}{m} \right)^{1/2} \| x^1 - x^0 \|; \tag{6.9}$$

$$f(x^{k+1}) - f(x^0) \leqslant \frac{M}{m} \left(\frac{M-m}{M+m} \right)^{2(k+1)} (f(x^1) - f(x^0)). \tag{6.10}$$

Таким образом, метод Коши сходится со скоростью геометрической прогрессии со знаменателем $\gamma = (M-m)/(M+m)$. В качестве m и M можно взять наименьшее и, соответственно, наибольшее собственные значения матрицы $\nabla^2 f(x)$, так что скорость сходимости, очевидно, в значительной мере зависит от отношения наибольшего собственного значения гессиана функции f к наименьшему. Знаменатель прогрессии γ лишь тогда будет достаточно малым, когда M и m незначительно отличаются друг от друга.

Эту ситуацию поясняет следующий пример.

Пример 6.6. Для $f(x) = (x_1^2/a^2 + x_2^2/b^2)/2$, $a > b > 0$, имеем $m = 1/a^2$, $M = 1/b^2$. В методе Коши каждые два следующие друг за другом направления движения взаимно ортогональны. На рис. 6.5, *а – г* приведены несколько итераций в зависимости от отношения $M/m = a^2/b^2$ и от угла между направлением движения на первом шаге и главной осью.

Покажем линейную скорость сходимости последовательности $\{f(x^k)\}$ к значению $f(x^0)$ для частного случая $f(x) = \frac{1}{2} x'Qx + a'x$, где Q – симметрическая положительно определенная матрица размера $n \times n$, $a \in \mathbf{R}^n$.

Для произвольного $q \in \mathbf{R}^n$ справедливо

$$f(x^k - r_k q) = f(x^k) + \frac{1}{2} r_k^2 q'Qq - r_k q' \nabla f(x^k). \tag{6.11}$$

Длина шага r_k из условия $\min\limits_{r \geqslant 0} f(x^k - rq) = f(x^k - r_k q)$ получается явно:

$$r_k = \frac{q' \nabla f(x^k)}{q'Qq}. \tag{6.12}$$

Далее, в силу $\nabla f(x^k) = Qx^k + a$ имеем

$$f(x^k) = \frac{1}{2} \nabla f(x^k)' Q^{-1} \nabla f(x^k) + \frac{1}{2} a' Q^{-1} a. \tag{6.13}$$

Из (6.11) – (6.13) при $f(x^0) = \frac{1}{2} a' Q^{-1} a$, $\nabla f(x^k) = q$ и $x^{k+1} = x^k - r_k \nabla f(x^k)$ получим

$$f(x^{k+1}) - f(x^0) = \left(1 - \frac{(q'q)^2}{(q'Qq)(q'Q^{-1}q)} \right) (f(x^k) - f(x^0)). \tag{6.14}$$

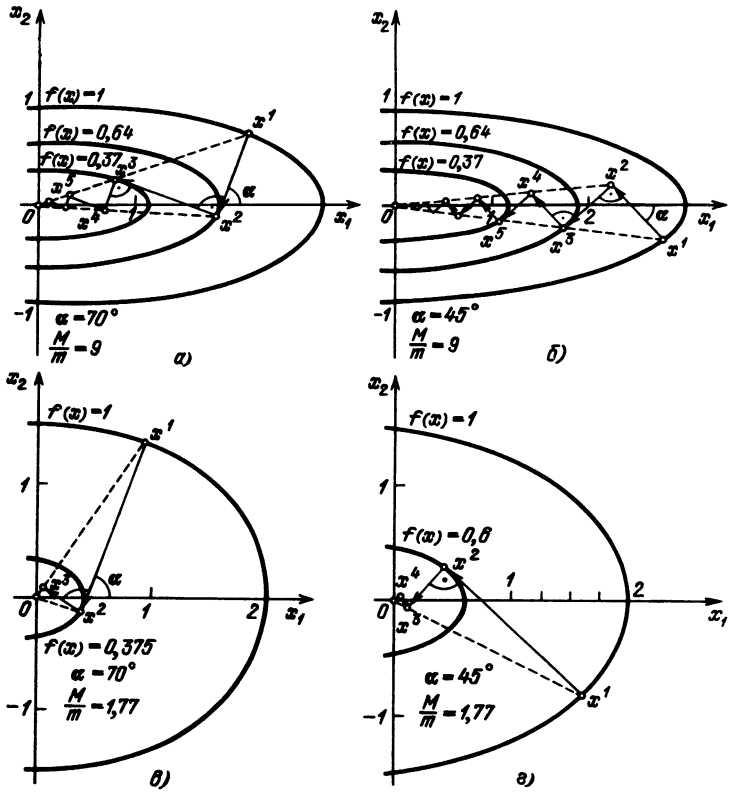


Рис. 6.5.

Для наибольшего и наименьшего собственных значений M и m положительно определенной матрицы размера $n \times n$ имеет место оценка

$$\frac{(q'q)^2}{(q'Qq)(q'Q^{-1}q)} \geqslant \frac{4mM}{(m+M)^2}, \tag{6.15}$$

что вместе с (6.14) и дает (6.10).

Градиентные методы с другим способом выбора длины шага исследуются, например, в работах [69, 101].

Методы Ньютона. Пусть функция $f\colon \mathbf{R}^n \to \mathbf{R}$ дважды непрерывно дифференцируема. Тогда для произвольного $x^0 \in \mathbf{R}^n$ справедливо разложение

$$f(x) = f(x^0) + (x - x^0)'\nabla f(x^0) + \frac{1}{2}(x - x^0)'\nabla^2 f(x^0)(x - x^0) +$$

$$+ o(\| x - x^0 \|^2) \quad \forall x \in \mathbf{R}^n.$$

Если f аппроксимировать квадратичной функцией

$$f^0(x) = f(x^0) + (x - x^0)'\nabla f(x^0) + \frac{1}{2}(x - x^0)'\nabla^2 f(x^0)(x - x^0),$$

то необходимое для наличия минимума f^0 условие будет иметь вид

$$\nabla f(x^0) + \nabla^2 f(x^0)(x - x^0) = 0. \tag{6.16}$$

Пусть матрица $\nabla^2 f(x^0)$ невырождена. Тогда решением (6.16) будет

$$x = x^0 - (\nabla^2 f(x^0))^{-1} \nabla f(x^0).$$

Отсюда выводится следующее рекуррентное правило, лежащее в основе метода Ньютона:

$$x^{k+1} = x^k - r_k (\nabla^2 f(x^k))^{-1} \nabla f(x^k) \qquad \forall k \geqslant 1. \tag{6.17}$$

Если длину шага r_k выбирать из условия

$$f(x^k - r_k (\nabla^2 f(x^k))^{-1} \nabla f(x^k)) = \min_{r \geqslant 0} f(x^k - r(\nabla^2 f(x^k))^{-1} \nabla f(x^k)), \tag{6.18}$$

то можно получить следующее утверждение относительно свойств сходимости метода Ньютона ([98, 101]).

Т е о р е м а 6.2. 1) *Пусть функция* f: $\mathbf{R}^n \to \mathbf{R}$ *дважды непрерывно дифференцируема и выполнено условие* (6.8). *Тогда последовательность* $\{x^k\}$, *построенная согласно* (6.17), (6.18), *сверхлинейно сходится к решению* x^0.

2) *Если, кроме того, в матричной норме*

$$\| \nabla^2 f(x) \| = \left[\sum_{i,j=1}^{n} \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right)^2 \right]^{1/2}$$

выполняется неравенство

$$\| \nabla^2 f(x) - \nabla^2 f(y) \| \leqslant K_2 \| x - y \| \qquad \forall x, y \in \mathbf{R}^n$$

для некоторого $K_2 > 0$, *то* $\{x^k\}$ *квадратично сходится к* x^0:

$$\| x^{k+1} - x^0 \| \leqslant \left(\frac{M}{m} \right)^{1/2} \frac{K_2}{m} \| x^k - x^0 \|^2.$$

В случае $r_k = r = 1$ *сходимость последовательности* $\{x^k\}$ *из* (6.17) *к решению* x^0 *можно обеспечить лишь в достаточно малой окрестности точки* x^0.

В методе Ньютона преимуществу, состоящему в более высокой по сравнению с градиентными методами скорости сходимости, противостоит недостаток, заключающийся в больших вычислительных трудностях на каждом шаге. Чтобы сочетать преимущества градиентных методов (применение производных лишь первого порядка) с преимуществами методов Ньютона (высокая скорость сходимости), разработано большое число методов, основную идею которых мы поясним ниже. А именно, мы рассмотрим классы квазиньютоновских методов и методов сопряженных направлений.

Квазиньютоновские методы. Эти методы основываются на аппроксимации матрицы $[\nabla^2 f(x)]^{-1}$, при этом используются лишь первые производные функции f. Итерации (6.17) метода Ньютона заменяются на

$$x^{k+1} = x^k - r_k F_k \nabla f(x^k) \qquad \forall k \geqslant 1, \tag{6.19}$$

где F_k – симметрическая матрица размера $n \times n$. При $F_k = E$ получается градиентный метод, а при $F_k = [\nabla^2 f(x^k)]^{-1}$ – метод Ньютона. Необходимое условие для того, чтобы определяемое через $-F_k \nabla f(x^k)$ направление было при $\nabla f(x^k) \neq 0$ направлением спуска, имеет вид

$$\nabla f(x^k)' F_k \nabla f(x^k) > 0 \quad \forall x^k \in X \cap \{x: \nabla f(x) \neq 0\}.$$

Это условие выполняется, если матрица F_k положительно определена.

Рассмотрим в качестве примера снова минимизацию функции $f(x) = \frac{1}{2} x'Qx + a'x$, где Q – симметрическая положительно определенная матрица. Длину шага r_k будем выбирать из соотношения

$$\min_{r \geq 0} f(x^k - rF_k p) = f(x^k - r_k F_k p), \qquad p = \nabla f(x^k).$$

Величина r_k определяется явно:

$$r_k = \frac{p' F_k p}{p' F_k Q F_k p}. \tag{6.20}$$

Из соотношений (6.17) (при $q = F_k p$), (6.13), (6.20) получаем

$$f(x^k - r_k F_k p) - f(x^0) = \left[1 - \frac{(p' F_k p)^2}{(p' F_k Q F_k p)(p' Q^{-1} p)}\right] (f(x^k) - f(x^0)). \tag{6.21}$$

Для симметрической положительно определенной матрицы F существует положительно определенная матрица H такая, что $H'H = F$. Положим, $s = Hp$, $H'QH = R$. Тогда $R^{-1} = H^{-1} Q^{-1} H'^{-1}$ и, учитывая (6.19), имеем

$$f(x^{k+1}) - f(x^0) = \left[1 - \frac{(s's)^2}{(s'Rs)(s'R^{-1}s)}\right] (f(x^k) - f(x^0)).$$

Отсюда согласно (6.15) получаем

$$f(x^{k+1}) - f(x^0) \leq \left(\frac{M' - m'}{M' + m'}\right)^2 (f(x^k) - f(x^0)), \tag{6.22}$$

где m' и M' – соответственно наименьшее и наибольшее собственные значения матрицы R.

В силу $FQ = HRH^{-1}$ матрица FQ подобна матрице R, и поэтому собственные значения FQ и R совпадают. Следовательно, выбирая подходящим образом матрицу F, можно существенно увеличить скорость сходимости по сравнению с градиентным методом. В соответствии со свойствами методов Ньютона надо стремиться, чтобы F аппроксимировала матрицу Q^{-1} (а в случае неквадратичной функции – матрицу $[\nabla^2 f(x^k)]^{-1}$).

Для того чтобы приблизить $[\nabla^2 f(x^k)]^{-1}$, в квазиньютоновском методе Давидона – Флетчера – Пауэлла ([56, 69]) применяют градиенты, вычисляемые в ходе итерационного процесса.

Методы сопряженных направлений. В этих методах тоже используются лишь градиенты функции f, при этом в основе построения методов вновь лежат свойства квадратичных функций.

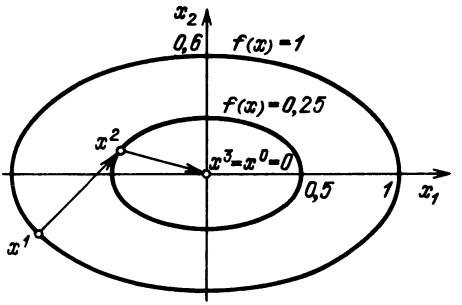


Рис. 6.6.

Векторы q^i и q^j $(i, j \in \{1; \ldots, n\})$ называются сопряженными относительно симметрической матрицы Q, если

$$(q^i)'Qq^j = 0 \qquad \forall i \neq j.$$

Если матрица Q положительно определена и векторы $q^1, \ldots, q^n$ попарно сопряжены относительно Q, то $q^1, \ldots, q^n$ линейно независимы.

Для минимизации функции $f(x) = x'Qx + a'x$ с положительно определенной матрицей Q применим итерационную процедуру

$$x^{k+1} = x^k - r_k q^k, \qquad k = 1, \ldots, n,$$

где векторы $q^1, \ldots, q^n$ сопряжены относительно матрицы Q, а число r_k выбрано из условия

$$\min_{r \geqslant 0} f(x^k - rq^k) = f(x^k - r_k q^k), \qquad k = 1, \ldots, n.$$

Покажем, что такой алгоритм определяет точку минимума квадратичной функции не более чем за n шагов.

Так как векторы q^k $(k = 1, \ldots, n)$ линейно независимы, то при некоторых λ_k справедливо равенство

$$x^{n+1} = x^1 + \sum_{k=1}^{n} \lambda_k q^k.$$

По определению метода

$$x^k = x^1 - \sum_{i=1}^{k-1} r_i q^i.$$

Отсюда, с одной стороны, следует

$$(q^k)'Qx^{n+1} = (q^k)'Qx^1 + \lambda_k (q^k)'Qq^k,$$

а с другой –

$$(q^k)'Qx^k = (q^k)'Qx^1.$$

Таким, образом, выполняется равенство

$$(q^k)'Qx^{n+1} - (q^k)'Qx^k = \lambda_k (q^k)'Qq^k.$$

Поскольку $\nabla f(x^k) = Qx^k + a$, $\nabla f(x^{n+1}) = Qx^{n+1} + a = 0$, отсюда имеем

$$-(q^k)'\nabla f(x^k) = \lambda_k (q^k)'Qq^k,$$

и, следовательно, $\lambda_k = -r_k$ для $k = 1, \ldots, n$.

На рис. 6.6 приведены итерации метода сопряженных направлений для функции из примера 6.6 при $a = 1$и $b = 0{,}6$. Методы сопряженных направлений разрабатывались также для неквадратичных функций и имеют большие преимущества перед градиентными методами ([56, 69]). Один из хорошо зарекомендовавших себя методов принадлежит Флетчеру и Ривсу (см. [40, 56, 69]).

6.3. Методы прямого поиска

Обсуждая проблему нахождения решения задачи

$$\min \{ f(x): \ x \in G \}$$

с помощью методов прямого поиска, мы ограничимся важным случаем $G = [a, b] \subseteq \mathbf{R}$. Более подробное рассмотрение этих методов можно найти, например, в работе [10].

В тесной связи с методами прямого поиска находятся класс унимодальных функций и некоторые обобщения этих функций (см. [10, 12]).

Как уже упоминалось в § 6.2, при выборе одной точки $x^1 \in G$ и определении $f(x^1)$ не извлекается никакой информации о собственном подмножестве множества G, в котором следует искать решение исходной задачи. Напротив, любой набор $(\xi^1, \ldots, \xi^k)$, $k \geqslant 2$, попарно различных точек из $\operatorname{int} G$ и значения функции f в этих точках при определенных предположениях дают интервал длины меньшей, чем $b - a$, который содержит решение x^0 поставленной задачи.

Функция f: $X \to \mathbf{R}$ называется *квазиунимодальной* на множестве $[a, b] \subseteq X \subseteq \mathbf{R}$, если найдется (произвольный) интервал $I^* \subseteq [a, b]$ такой, что функция f строго убывает на $[a, c]$, равна константе, не превосходящей $\min\{f(c), f(d)\}$, на I^* и строго возрастает на $[d, b]$, где $c = \inf\limits_{x \in I^*} x$, $d = \sup\limits_{x \in I^*} x$. Ясно, что I^* является также множеством решений задачи

$$\min \{ f(x): \ x \in [a, b] \} . \tag{6.23}$$

В частности, если $c = d$, то $I^* = \{x^0\}$ и $x^0 = c = d$. Такие функции называются *унимодальными**) на множестве $[a, b]$. Свойство унимодальности характеризуется существованием единственного локального минимума на $[a, b]$. У квазиунимодальных функций возможны также частные случаи $a = c$, $d = b$ и $I^* = [a, b]$. Поэтому любая функция, постоянная на $[a, b]$, является там и квазиунимодальной. Другие примеры квазиунимодальных функций представлены на рис. 6.7. Соотношения между квазиунимодальными и унимодальными на сегменте $[a, b]$ функциями и рассмотренными в § 2.3 обобщениями выпуклых функций описаны в следующем утверждении.

Теорема 6.3. 1) *Если функция* f: $[a, b] \to \mathbf{R}$ *квазиунимодальна на множестве* $[a, b]$, *то* f *почти выпукла на* $[a, b]$.

2) *Если* f *унимодальна на* $[a, b]$, *то* f *строго явно квазивыпукла на* $[a, b]$.

Доказательство. 1. Если f квазиунимодальна на $[a, b]$, то справедливы соотношения

$$\left.\begin{array}{l} x^1, x^2 \in [a, b], \quad \lambda \in (0, 1) \\ f(x^2) < f(x^1) \end{array}\right\} \Rightarrow f(\lambda x^1 + (1 - \lambda) x^2) < f(x^1),$$

$$\left.\begin{array}{l} x^1, x^2 \in [a, b], \quad \lambda \in [0, 1) \\ f(x^1) = f(x^2) \end{array}\right\} \Rightarrow f(\lambda x^1 + (1 - \lambda) x^2) \leqslant f(x^1) . \tag{6.24}$$

*) Это название заимствовано из свойств функции плотности некоторого непрерывного распределения вероятностей.

Тем самым имеем

$$\left.\begin{array}{l} x^1, x^2 \in [a, b], \quad \lambda \in [0, 1] \\ f(x^2) \leqslant f(x^1) \end{array}\right\} \Rightarrow f(\lambda x^1 + (1-\lambda)x^2) \leqslant f(x^1).$$

Но это – определение явной квазивыпуклости функции f на множестве $[a, b]$. Отсюда с учетом (6.24) следует, что f почти выпукла на $[a, b]$.

2. Допустим, что f не является строго явно квазивыпуклой на $[a, b]$. Тогда найдутся $x^1, x^2 \in [a, b]$ (без ограничения общности, пусть $x^1 < x^2$) и число $\lambda^* \in (0, 1)$ такие, что $f(x^2) \leqslant f(x^1)$ и

$$f(\lambda^* x^1 + (1-\lambda^*)x^2) \geqslant f(x^1).$$

Однако функция f не унимодальна на сегменте $[x^1, x^2]$, потому что для $\tilde{x} < \lambda^* x^1 + (1-\lambda^*)x^2$ она не возрастает строго на $[\tilde{x}, x^2]$ и для $\tilde{x} \geqslant \lambda^* x^1 + (1-\lambda^*)x^2$ не убывает строго на $[x^1, \tilde{x}]$. Так как любая унимодальная на $[a, b]$ функция будет унимодальной и на всех замкнутых подынтервалах из $[a, b]$ (см., например, [11], лемма 1), т.е. и на $[x^1, x^2]$, то получается противоречие. Тем самым первоначальное допущение неверно.

З а м е ч а н и е. Можно показать, что в случае существования локального минимума функции f на сегменте $[a, b]$ утверждения теоремы 6.3 обратимы. В частности, это справедливо для функций, которые полунепрерывны снизу на $[a, b]$. Кроме того, в этом случае понятия "почти выпуклая" и "строго квазивыпуклая" совпадают.

Связь между строго явно квазивыпуклыми (SEQX), почти выпуклыми (FKX), унимодальными (U) и квазиунимодальными (QU) функциями f: $[a, b] \to \mathbf{R}$ показана на рис. 6.8.

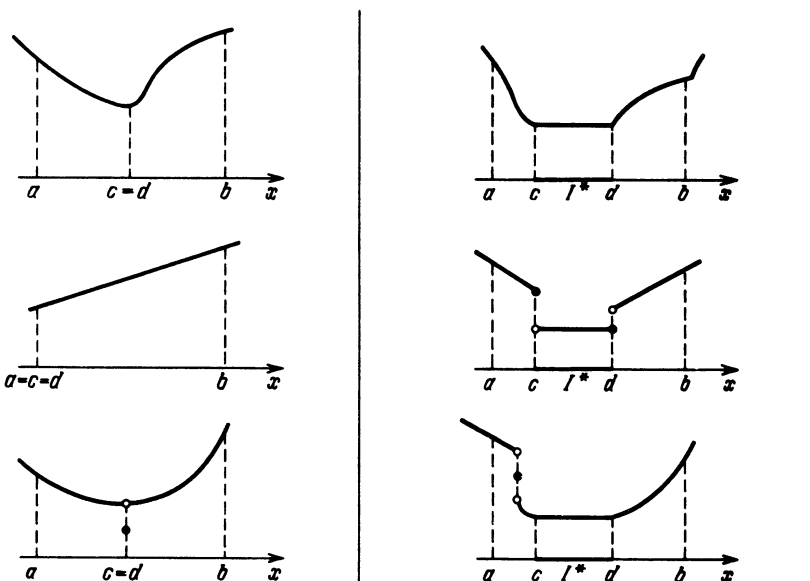


Рис. 6.7.

Рис. 6.8.

Обсудим теперь упоминавшуюся выше проблему поиска подынтервала из $[a, b]$, содержащего решение x^0 задачи (6.23).

Теорема 6.4. *Пусть функция* f: $[a, b] \to \mathbf{R}$ *квазиунимодальна на* $[a, b]$, $a = \xi^0 < \xi^1 < \ldots < \xi^k < \xi^{k+1} = b$ *и*

$$f(\xi^l) = \min_{i \in \{1, \ldots, k\}} \{f(\xi^i)\}, \quad l \in \{1, \ldots, k\}. \tag{6.25}$$

Тогда существует решение x^0 *задачи* (6.23) *и*

$$x^0 \in (\xi^{l-1}, \xi^{l+1}). \tag{6.26}$$

Доказательство. Если допустить, что

$$f(x) > f(x^0) \qquad \forall x \in (\xi^{l-1}, \xi^{l+1}), \qquad \forall x^0 \in I^*,$$

то из квазиунимодальности функции f на $[a, b]$ следует, что f или строго убывает на (ξ^{l-1}, ξ^{l+1}) или строго возрастает. В обоих случаях получаем противоречие с (6.25). Следовательно, выполняется (6.26).

Пусть после k вычислений значений функции в методе прямого поиска для решения задачи (6.23) величина D_k будет наименьшей верхней оценкой для длины интервала (ξ^{l-1}, ξ^{l+1}), в котором заведомо содержится решение x^0 задачи (6.23). Очевидно, что D_k будет зависеть от выбора k – набора $(\xi^1, \ldots, \xi^k)$. Для сравнения эффективности этих методов при решении задачи (6.23) используем отношение

$$\widetilde{L}_k = D_k/(b - a).$$

Одновременное сравнение значений функции. При одновременном сравнении значений функции отношение $\widetilde{L}_k$ равно инфимуму чисел

$$L_k = \max_{l \in \{1, \ldots, k\}} \frac{\xi^{l+1} - \xi^{l-1}}{b - a}. \tag{6.27}$$

Если число $k > 1$ нечетное, то значение L_k будет минимальным, когда мы произведем разбиение интервала $[a, b]$ на равные промежутки. При этом (см. рис. 6.9 для $k = 3$)

$$\widetilde{L}_k = \frac{\hat{\xi}^{l+1} - \hat{\xi}^{l-1}}{b - a} = \frac{2}{k+1} \leqslant L_k$$

для всех $(\xi^1, \ldots, \xi^k)$ таких, что $a < \xi^1 < \ldots < \xi^k < b$.

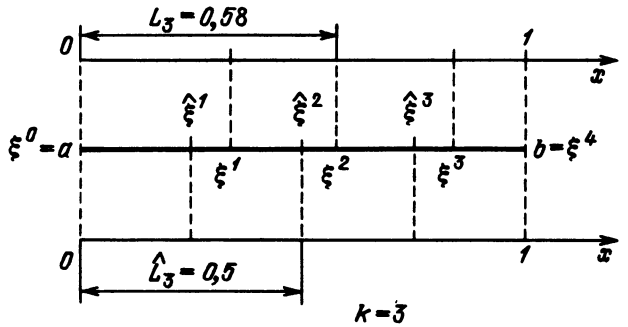


Рис. 6.9.

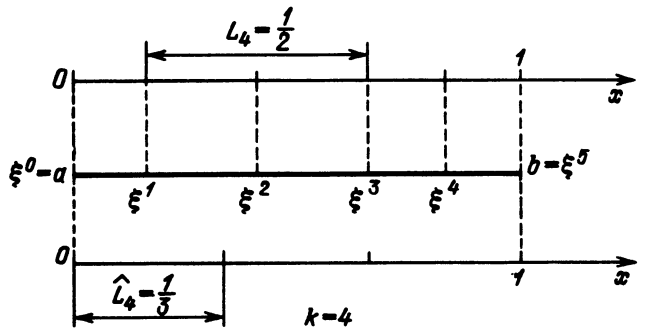


Рис. 6.10.

Если число k четное, то для величин L_k может быть указан лишь инфимум, который не достигается ни на одном k-наборе $(\xi^1, \ldots, \xi^k)$ со свойствами, приведенными в теореме 6.4. Действительно, в противном случае точки ξ^{2i-1} и ξ^{2i} для $i = 1, \ldots, k/2$ должны были бы совпадать, что противоречит названным свойствам k-набора (см. рис. 6.10 для $k = 4$). Итак, $2/(k+2) < L_k$ для всех $(\xi^1, \ldots, \xi^k)$ таких, что $a < \xi^1 < \ldots < \xi^k < b$.

Однако в цифровых ЭВМ для каждого представимого в этой машине числа $\bar{x}$ существует действительное число $\delta = \delta(\bar{x})$ такое, что для всех представимых в машине чисел $\bar{y} \neq \bar{x}$ выполняется $|\bar{x} - \bar{y}| \geqslant \delta$*). Поэтому при четном k для L_k может быть указано хотя бы одно отношение $\widehat{L}_k^{\delta}$, δ-минимальное в том смысле, что

$$|\xi^i - \xi^j| = \delta \quad \forall\, i, j \in \{0, 1, \ldots, k+1\}, \quad i \neq j. \tag{6.28}$$

---

*) Пусть, например, применяется представление числа с плавающей запятой $\bar{x} = mB^e$ с основанием B (скажем, 2 или 10), мантиссой m, $B^{-1} \leqslant |m| < 1$, целочисленным показателем e, и пусть m как B-ичная дробь имеет p знаков. Тогда значение последней в мантиссе цифры, в которой число $\bar{y} \neq \bar{x}$ с тем же показателем e должно минимально отличаться от $\bar{x}$, равно B^{-p}. Таким образом, в этом случае $\delta = B^{e-p}$.

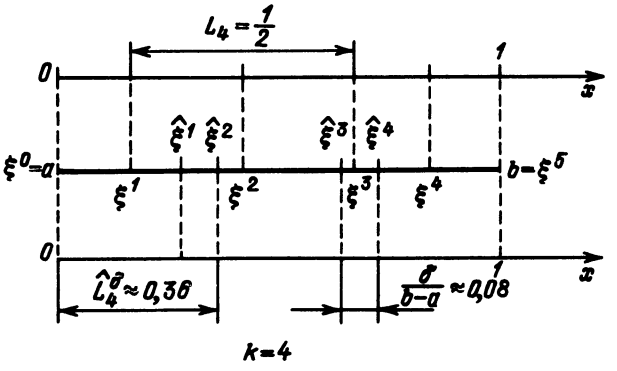


Рис. 6.11.

Следовательно, для любого k-набора $(\xi^1, \ldots, \xi^k)$, где $a = \xi^0 < \xi^1 < \ldots < \xi^k < \xi^{k+1} = b$, k четно, имеем неравенство

$$\frac{\xi^{l+1} - \xi^{l-1}}{b-a} \leqslant L_k \qquad \forall l \in \{1, \ldots, k\} .$$

Тем самым для того же k-набора получаем

$$\frac{\xi^k - \xi^0}{b-a} \leqslant \frac{k}{2} L_k$$

и, в силу (6.28),

$$1 = \frac{\xi^{k+1} - \xi^0}{b-a} = \frac{\xi^k - \xi^0}{b-a} + \frac{\xi^{k+1} - \xi^{k-1}}{b-a} - \frac{\xi^k - \xi^{k-1}}{b-a} \leqslant$$

$$\leqslant \frac{k+2}{2} L_k - \frac{\delta}{b-a}$$

Отсюда при заданном $\delta > 0$ следует

$$\widetilde{L}_k^\delta = \frac{2}{k+2} + \frac{2\delta}{(k+2)(b-a)} \leqslant L_k .$$

Соответствующий этому δ-минимальному отношению $\widetilde{L}_k^\delta$ k-набор $(\hat{\xi}^1, \ldots, \hat{\xi}^k)$ имеет вид (см. рис. 6.11 для $k = 4$)

$$\hat{\xi}^i = \begin{cases} a + \dfrac{i}{2} \widetilde{L}_k^\delta (b-a), & \text{если } i \text{ четное}, \\ b - \dfrac{k-i+1}{2} \widetilde{L}_k^\delta (b-a), & \text{если } i \text{ нечетное}. \end{cases}$$

Однако более эффективным для $k \geqslant 3$ оказывается подход, в котором определение ξ^i $(i > 1)$ зависит от результатов вычисления значений функции $f(\xi^j)$.

Последовательное сравнение значений функции. Здесь задается фиксированное число сравниваемых на каждом этапе значений функции, которые

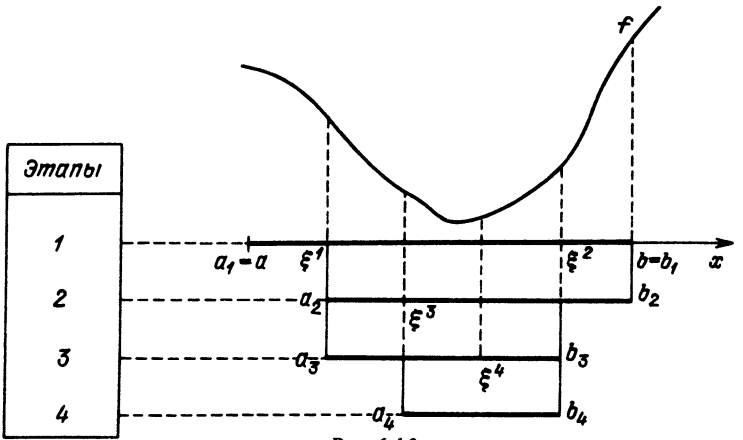


Рис. 6.12.

образуют так называемый блок, и точка с наименьшим значением функции одного этапа применяется и на следующем этапе. На рис. 6.12 мы поясняем такой последовательный метод для важного случая, когда одному блоку принадлежат два значения функции.

На первом этапе сравниваются $f(\xi^1)$ и $f(\xi^2)$. Так как $f(\xi^1) > f(\xi^2)$, то у квазиунимодальной на сегменте $[a, b]$ функции f ни одна точка из I^* не лежит в $[a, \xi^1)$. Этот подынтервал исключается из дальнейшего рассмотрения, остается $[a_2, b_2] = [\xi^1, b]$; точка $\xi^2 \in (a_2, b_2]$ переносится на второй этап. На втором этапе уже найденное значение $f(\xi^2)$ сравнивается с $f(\xi^3)$, $\xi^3 \in (a_2, b_2)$. В силу $f(\xi^2) > f(\xi^3)$ интервал (ξ^2, b_2) исключается, остается $[a_3, b_3] = [a_2, \xi^2]$; точка $\xi^3 \in (a_3, b_3)$ переносится на третий этап. На третьем этапе $f(\xi^3)$ сравнивается с $f(\xi^4)$, $\xi^4 \in (\xi^1, \xi^2)$. Так как $f(\xi^3) > f(\xi^4)$, то интервал $[a_3, \xi^3)$ исключается, остается $[a_4, b_4] = [\xi^3, b_3]$, а точка $\xi^4 \in (a_4, b_4)$ переносится на четвертый этап и т.д. При таком подходе возникают пары последовательностей с перечисленными ниже свойствами.

Пара последовательностей $(\{a_\nu\}, \{b_\nu\})$, $a_\nu, b_\nu \in [a, b] \subset \mathbf{R}$, называется порожденной последовательным сравнением значений функции в двухточечных блоках, если $a_1 = a$, $b_1 = b$ и имеются последовательности $\{\xi^{\nu 1}\}$, $\{\xi^{\nu 2}\}$, для которых

$$1)\ a_\nu = \xi^{\nu 0} < \xi^{\nu 1} < \xi^{\nu 2} < \xi^{\nu 3} = b_\nu \qquad \forall \nu \in \mathbf{N}; \tag{6.29а}$$

$$2)\ a_{\nu+1} = \xi^{\nu, j-1}, \qquad b_{\nu+1} = \xi^{\nu, j+1}, \tag{6.29б}$$

где $j \in \{1, 2\}$ – наименьший номер, для которого

$$f(\xi^{\nu j}) = \min\{f(\xi^{\nu 1}),\ f(\xi^{\nu 2})\} \qquad \forall \nu \in \mathbf{N};$$

$$3)\ \xi^{\nu+1, i_j} = \xi^{\nu j},\ \xi^{\nu+1, i_l} = \xi^{\nu j}, \quad \xi^{\nu+1, i_l} \in (a_{\nu+1}, b_{\nu+1}), \tag{6.30}$$

$$i_j \neq i_l, \quad i_j, i_l \in \{1, 2\} \quad \forall \nu \in \mathbf{N}.$$

Теореме 6.4 соответствует следующее утверждение.

Теорема 6.5. *Пусть функция* f: $[a, b] \to \mathbf{R}$ *квазиунимодальна на сегменте* $[a, b]$ *и* $(\{a_\nu\}, \{b_\nu\})$, *где* $a_\nu, b_\nu \in [a, b]$, – *порожденная последо-*

вательным сравнением значений функции пара последовательностей. Тогда существует решение x^0 задачи (6.23) *и* $x^0 \in [a_\nu, b_\nu]\ \forall \nu \in \mathbf{N}$.

Доказательство. Предположим противное. Пусть для всякого решения $x^0 \in I^*$ существует номер ν^* такой, что $x^0 \notin [a_{\nu^*}, b_{\nu^*}]$. Не ограничивая общности, можно допустить, что ν^* минимален в том смысле, что $x^0 \in [a_\nu, b_\nu]\ \forall \nu < \nu^*$.

В частности, $x^0 \in [a_{\nu^*-1}, b_{\nu^*-1}]$. В силу $a_{\nu^*-1} \leqslant a_{\nu^*} < b_{\nu^*} \leqslant b_{\nu^*-1}$ и (6.29б) следует различать два случая.

Случай 1: $j=1$. Тогда согласно (6.29а) и (6.29б) $a_{\nu^*} = a_{\nu^*-1} = \xi^{\nu^*-1,0}$, $b_{\nu^*} = \xi^{\nu^*-1,2}$ и $x^0 \in (b_{\nu^*}, b_{\nu^*-1})$. С одной стороны, из (6.29б) следует $f(x^0) \leqslant f(\xi^{\nu^*-1,1}) \leqslant f(\xi^{\nu^*-1,2})$, а с другой – $\xi^{\nu^*-1,1} < \xi^{\nu^*-1,2} < x^0$. Однако эти соотношения несовместимы с квазиунимодальностью функции f на $[a, b]$.

Случай 2: $j=2$. Тогда $b_{\nu^*} = b_{\nu^*-1} = \xi^{\nu^*-1,3}$, $a_{\nu^*} = \xi^{\nu^*-1,1}$ и имеет место $x^0 \in [a_{\nu^*-1}, a_{\nu^*})$. Из (6.29б) следует $f(x^0) \leqslant f(\xi^{\nu^*-1,2}) \leqslant f(\xi^{\nu^*-1,1})$. Одновременно выполняются неравенства $x^0 \leqslant \xi^{\nu^*-1,1} \leqslant \xi^{\nu^*-1,2}$, что тоже противоречит заданным предположениям.

Замечание. Пара последовательностей $(\{a_\nu\}, \{b_\nu\})$ не обязательно образует последовательность вложенных интервалов, поскольку длина интервала $(b_\nu - a_\nu)$ при $\nu \to \infty$ не стремится, вообще говоря, к нулю.

Далее мы изложим основные идеи построения (конечной) пары последовательностей $(\{a_\nu\}, \{b_\nu\})$, $\nu = 1, \ldots, k$, с δ-минимальным отношением $\tilde{L}_k^\delta$.

Пусть после $\nu \geqslant 2$ вычислений значений функции остался (рис. 6.13) интервал $[\xi^{\nu 0}, \xi^{\nu 3}]$ длины

$$D_\nu = (\xi^{\nu 3} - \xi^{\nu 0}). \tag{6.31}$$

Так как каждый из интервалов $[\xi^{\nu 0}, \xi^{\nu 2}]$, $[\xi^{\nu 1}, \xi^{\nu 3}]$ может содержать точку x^0, мы выберем длины интервалов $(\xi^{\nu 2} - \xi^{\nu 0})$ и $(\xi^{\nu 3} - \xi^{\nu 1})$ равными $D_{\nu+1} = \xi^{\nu 3} - \xi^{\nu 1} = \xi^{\nu 2} - \xi^{\nu 0}$ (это оптимальный выбор $D_{\nu+1}$). Допустим, что сегмент $[\xi^{\nu 1}, \xi^{\nu 3}]$ содержит x^0. Согласно (6.30) имеем тогда $\xi^{\nu+1,1} = \xi^{\nu 2}$. Новую итерационную точку $\xi^{\nu+1,2}$ выберем так, чтобы длины

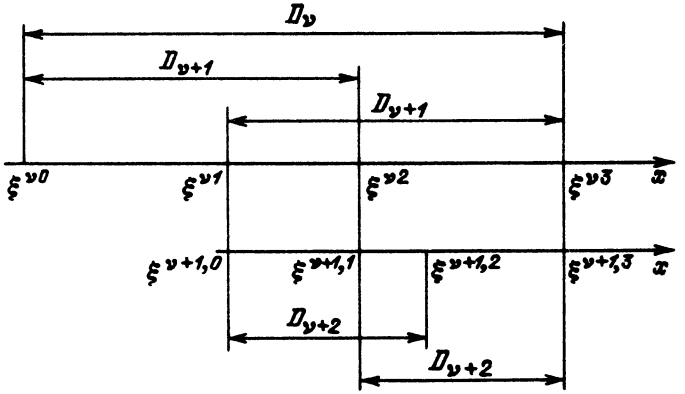


Рис. 6.13.

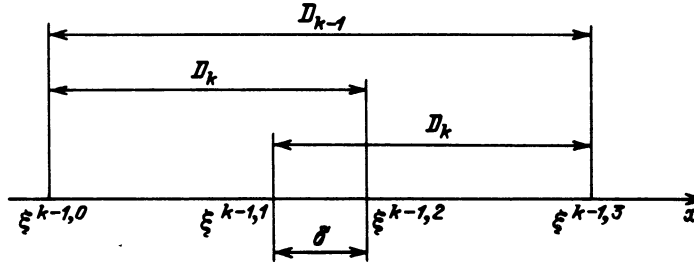


Рис. 6.14.

интервалов $(\xi^{\nu+1,2} - \xi^{\nu+1,0})$ и $(\xi^{\nu+1,3} - \xi^{\nu+1,1})$ снова были равными. Очевидно, что между длинами интервалов выполняется соотношение (см. рис. 6.13)

$$D_\nu = D_{\nu+1} + D_{\nu+2}, \quad \nu = 1, \ldots, k-2. \tag{6.32}$$

Потребуем, чтобы расстояние между итерационными точками $\xi^{k-1,1}$ и $\xi^{k-1,2}$ после $k \geqslant 2$ вычислений значений функции равнялось заданному числу $\delta > 0$, так что будет справедливо соотношение (рис. 6. 14)

$$D_{k-1} = 2D_k - \delta. \tag{6.33}$$

Следующая теорема утверждает, что существует итерационная последовательность $\xi^{\nu,i}$, $i = 0, \ldots, 3$, $\nu = 1, \ldots, k-1$, для которой выполняются соотношения (6.32), (6.33) и, кроме того, $D_1 = b - a$.

Т е о р е м а 6.6. *Для длин интервалов* D_ν, $\nu = 1, \ldots, k$, *из* (6.32) *и* (6.33) *следует*

$$D_\nu = \frac{F_{k-\nu+2}}{F_{k+1}}(b-a) + \frac{F_{k-1}F_{k-\nu+2} - F_{k+1}F_{k-\nu}}{F_{k+1}}\delta, \tag{6.34}$$

где $F_0 = 0$, $F_1 = 1$, $F_n = F_{n-1} + F_{n-2}$ $\forall n \geqslant 2$ – *числа Фибоначчи.*

Д о к а з а т е л ь с т в о. Вначале с помощью математической индукции покажем, что

$$D_\nu = F_{k-\nu+2}D_k - F_{k-\nu}\delta, \quad \nu = 1, \ldots, k. \tag{6.35}$$

Это утверждение верно для $\nu = k$, потому что на самом деле

$$F_2 D_k - F_0\delta = D_k.$$

В силу (6.33) равенство (6.35) верно и для $\nu = k-1$:

$$D_{k-1} = F_3 D_k - F_1\delta = 2D_k - \delta.$$

Предположим теперь, что (6.35) верно для всех j таких, что $k \geqslant j \geqslant \nu \geqslant 1$. Тогда, применяя предположение индукции, для $\nu \leqslant k-2$ согласно (6.32) будем иметь

$$D_\nu = F_{k-\nu+1}D_k - F_{k-\nu-1}\delta + F_{k-\nu}D_k - F_{k-\nu-2}\delta.$$

Отсюда в силу рекурсивного определения чисел Фибоначчи вытекает соотношение (6.35). Если в (6.35) положить $\nu = k$, то

$$D_k = \frac{b - a + F_{k-1}\delta}{F_{k+1}} .$$

Из последнего равенства и (6.35) получаем (6.34).

З а м е ч а н и е. Приведем несколько первых чисел Фибоначчи: $F_2 = 1$, $F_3 = 2, F_4 = 3, F_5 = 5, F_6 = 8, F_7 = 13$ и далее: 21, 34, 55, 89, 144, ...

В частности, для $\nu = 2$ из (6.34) получим

$$D_2 = \frac{F_k}{F_{k+1}} (b - a) + \frac{F_{k-1}F_k - F_{k+1}F_{k-2}}{F_{k+1}} \delta =$$

$$= \frac{F_k}{F_{k+1}} (b - a) + \frac{F_{k+1}F_{k-1} - F_k^2}{F_{k+1}} \delta. \qquad (6.36)$$

Учитывая тождество

$$F_{n+1}^2 - F_n F_{n+2} = (-1)^n \quad \forall n \in \mathbf{N}, \qquad (6.37)$$

которое мы докажем по индукции, для $n = k - 1$ из (6.36) имеем

$$D_2 = \frac{F_k}{F_{k+1}} (b - a) + \frac{(-1)^k}{F_{k+1}} \delta. \qquad (6.38)$$

Докажем теперь (6.37). Для $n = 1$ это соотношение, очевидно, верно: $F_2^2 - F_1 F_3 = 1 - 1 \cdot 2 = (-1)^1$.

Предположим, что (6.37) верно для $n = i \geqslant 1$. Тогда из соотношения (6.37) при $n = i + 1$, предположения индукции и определения $F_n = F_{n-1} + F_{n-2}$ при $n = i + 2$ и $n = i + 3$ следует

$$F_{i+2}^2 - F_{i+1}F_{i+3} = F_{i+2}(F_i + F_{i+1}) - F_{i+1}(F_{i+1} + F_{i+2}) =$$

$$= F_{i+2}F_i - F_{i+1}^2 = -(-1)^i = (-1)^{i+1}.$$

С помощью формул (6.38) и (6.32), исходя из $D_1 = b - a$, можно вычислить рекурсивным образом все D_ν, $\nu = 1, 2, ..., k$.

Тем самым мы имеем возможность описать конкретную пару конечных последовательностей, порожденную последовательным сравнением значений функции в двухточечных блоках.

Пара конечных последовательностей $(\{a_\nu\}, \{b_\nu\})$, $\nu = 1, ..., k$, называется *парой последовательностей Фибоначчи* (F-парой), если

1) $a_1 = a$, $b_1 = b$, $\xi^{11} = b_1 - D_2$, $\xi^{12} = a_1 + D_2$;

2) $a_{\nu+1} = \begin{cases} a_\nu & \text{при } f(\xi^{\nu 1}) \leqslant f(\xi^{\nu 2}), \\ \xi^{\nu 1} & \text{в противном случае,} \end{cases}$ $\nu = 1, ..., k - 1;$

Таблица 6.2

ν	$D_{\nu+1}$	a_ν	$\xi^{\nu 1} = \begin{cases} b_\nu - D_\nu + 1 \\ \xi^{\nu-1,2} \end{cases}$	$f(\xi^{\nu 1})$	$\xi^{\nu 2} = \begin{cases} \xi^{\nu-1,1} \\ a_\nu + D_{\nu+1} \end{cases}$	$f(\xi^{\nu 2})$	b_ν	$D_\nu = b_\nu - a_\nu$
1	$\frac{5}{8}(b-a) - \frac{1}{8}\delta = 1{,}24$	−0,50	(1) 1,50 − 1,24 = 0,26	0,068	(2) −0,50 + 1,24 = 0,74	0,55	1,50	2,00
2	$\frac{3}{8}(b-a) + \frac{1}{8}\delta = 0{,}76$	−0,50	(3) 0,74 − 0,76 = −0,02	0,0004	0,26	0,068	0,74	1,24
3	$\frac{2}{8}(b-a) - \frac{2}{8}\delta = 0{,}48$	−0,50	(4) 0,26 − 0,48 = −0,22	0,049	−0,02	0,0004	0,26	0,48
4	$\frac{1}{8}(b-a) + \frac{3}{8}\delta = 0{,}29$	−0,22	−0,02	0,0004	(5) −0,22 + 0,29 = 0,07	0,0049	0,26	0,48
5		−0,22					0,07	0,29

3) $b_{\nu+1} = \begin{cases} \xi^{\nu 2} & \text{при } f(\xi^{\nu 1}) \leqslant f(\xi^{\nu 2}), \\ & \nu = 1, \ldots, k-1; \\ b_\nu & \text{в противном случае,} \end{cases}$

4) $\xi^{\nu+1,l} = \begin{cases} b_{\nu+1} - D_{\nu+2} & \text{при } a_\nu = a_{\nu+1}, \\ & \nu = 1, \ldots, k-2;\ l = 1, 2; \\ a_{\nu+1} + D_{\nu+2} & \text{при } b_\nu = b_{\nu+2}, \end{cases}$

$\xi^{\nu+1,l} \neq \xi^{\nu j}$, где $j \in \{1, 2\}$ – наименьший номер такой, что

$f(\xi^{\nu j}) = \min \{ f(\xi^{\nu 1}), f(\xi^{\nu 2}) \}$.

При этом D_ν определены согласно (6.34) или согласно (6.32) и (6.38). Порядок действий при построении F-пары поясняется на следующем примере.

Пример 6.7. Пусть $f(x) = x^2$, $a = -0{,}5$, $b = 1{,}5$, $\delta = 0{,}09$ и $k = 5$. Табл. 6.2 содержит числовые значения, найденные в соответствии с предыдущими определениями, а также имеющиеся между этими значениями связи. В качестве результата получаем

$x^0 \in [a_5, b_5] = [-0{,}22,\ 0{,}07]$, $f(x^0) \leqslant \min \{ f(\xi^{41}), f(\xi^{42}) \} = 0{,}0004$.

Рис. 6.15 поясняет ситуацию.

Как было показано в работах Оливера и Вильде [65], Авризля и Вильде [3] (в обобщенном виде с применением блоков из $2p$, $p \geqslant 1$, значений функции), пары последовательностей Фибоначчи определяют δ-минимальное отношение $\tilde{L}_k^\delta$ в том случае, когда номер k выбирается из условия

$$D_k \geqslant 2\delta. \tag{6.39}$$

Среди натуральных чисел k, которые удовлетворяют (6.39), имеется наибольшее число k_0. Очевидно, должно выполняться

$$2\delta \leqslant D_{k_0} < 3\delta, \tag{6.40}$$

поскольку в случае $D_{k_0} \geqslant 3\delta$ и $D_{k_0+1} = D_{k_0} - \delta$ получаем $D_{k_0+1} \geqslant 2\delta$, что противоречит свойству максимальности k_0.

Применяя (6.34) при $\nu = k_0$ и рекурсивное соотношение для чисел Фибоначчи, из неравенств (6.40) получаем

$$F_{k_0+2} \leqslant (b-a)/\delta < F_{k_0+3}. \tag{6.41}$$

В примере 6.7 число k_0 определено согласно этому неравенству.

Неравенство $D_{k_0} < 3\delta$ является одновременно оценкой для длины наименьшего определяемого интервала $[a_{k_0}, b_{k_0}]$, в котором должна находиться точка минимума функции f.

Мы укажем, однако, на то, что рекурсивное нахождение величин D_ν ($\nu = 3, 4, \ldots, k$), исходя из $D_1 = b - a$, D_2, где D_2 определяется в соответствии с (6.38), численно неустойчиво. А именно, если D_2 взято с ошибкой ϵ:

$$\bar{D}_2 = D_2 + \epsilon, \tag{6.42}$$

то

$$\bar{D}_\nu = D_\nu + (-1)^\nu F_{\nu-1}\epsilon, \quad \nu = 2, \ldots, k. \tag{6.43}$$

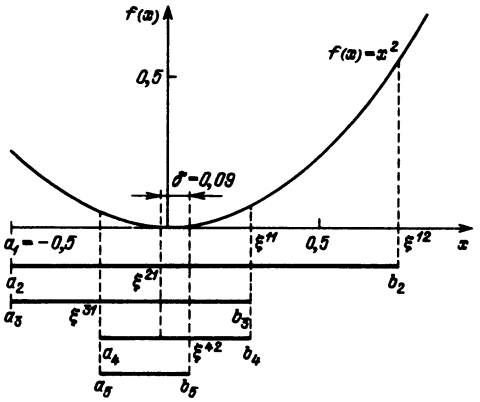


Рис. 6.15.

Докажем это. В силу (6.42) соотношение (6.43) верно для $\nu = 2$. Предположим, что оно верно и для $\nu < j \leqslant k$. Из (6.32) имеем

$$\bar{D}_j = \bar{D}_{j-2} - \bar{D}_{j-1},$$

откуда по предположению индукции следует

$$\bar{D}_j = D_{j-2} + (-1)^{j-2} F_{j-3}\epsilon - D_{j-1} - (-1)^{j-1} F_{j-2}\epsilon = D_j + (-1)^{j-2}(F_{j-3} + F_{j-2})\epsilon = D_j + (-1)^j F_{j-1}\epsilon.$$

В представлении (6.43) положительная последовательность $\{D_\nu\}$ убывает, в то время как модули вторых слагаемых становятся сколь угодно большими. Это является одной из причин, по которым при решении задачи (6.23) приходится, как правило, отказываться от применения F-пар, несмотря на некоторый выигрыш в эффективности. Кроме того, при этом речь идет не о методе решения, который сходится в смысле данного в § 6.1 определения, поскольку последовательность $\{x^\nu\}$ (например, $x^\nu = \xi^{\nu 1}$), вообще говоря, обрывается, не обязательно достигнув решения x^0.

Если пренебречь в (6.34) влиянием на D_ν второго слагаемого в предположении, что разыскиваемые при этом $\xi^{\nu j}$ $(\nu \in \mathbf{N}, j \in \{1, 2\})$ лежат во всюду плотной числовой области, то можно привести рекурсивное соотношение между длинами интервалов (обозначаемых через $\tilde{D}_\nu$), которое будет численно устойчиво.

С этой целью мы применим асимптотическое приближение для чисел Фибоначчи: из явного представления этих чисел

$$F_\nu = \frac{\tau^\nu - (-\tau)^{-\nu}}{\sqrt{5}}, \text{ где } \tau = \frac{1+\sqrt{5}}{2},$$

в силу $\tau > 1$ следует

$$F_\nu \approx \tau^\nu / \sqrt{5} \tag{6.44}$$

для достаточно большого ν.

Если мы положим $\tilde{D}_1 = D = b - a$ и в первом слагаемом в (6.34) вместо $F_{k-\nu+2}/F_{k+1}$ возьмем в соответствии с (6.44) приближенное значение $1/\tau^{\nu-1}$, то вместо (6.34) получим не зависящее от k и тем самым от δ выражение для длин интервалов:

$$\tilde{D}_\nu = \frac{1}{\tau^{\nu-1}}(b-a) = \frac{1}{\tau^{\nu-1}}\tilde{D}_1, \quad \nu \in \mathbf{N}. \tag{6.34'}$$

С помощью полной индукции легко доказать рекурсивное соотношение

$$\tilde{D}_{\nu+1} = \frac{1}{\tau}\tilde{D}_\nu, \quad \nu \in \mathbf{N}. \tag{6.45}$$

Нахождение $\tilde{D}_\nu$ в соответствии с (6.45), начиная с $\tilde{D}_1 = b - a$, численно устойчиво. В самом деле, пусть $\tilde{D}_2$ наделено ошибкой ϵ:

$$\bar{D}_2 = \tilde{D}_2 + \epsilon.$$

Тогда

$$\bar{D}_{\nu+1} = \tilde{D}_{\nu+1} + \epsilon/\tau^{\nu-1} \qquad \forall\, \nu \geqslant 1. \tag{6.46}$$

Доказывая (6.46), мы заметим вначале, что это соотношение верно для $\nu = 1$. Пусть оно верно для $\nu = k - 1$ $(k \geqslant 2)$. Из (6.45) имеем $\bar{D}_{k+1} = \frac{1}{\tau}\,\bar{D}_k$, откуда по предположению индукции следует

$$\bar{D}_{k+1} = \frac{1}{\tau}\left(\tilde{D}_k + \frac{\epsilon}{\tau^{k-2}}\right) = \frac{1}{\tau}\,\tilde{D}_k + \frac{\epsilon}{\tau^{k-1}} = \tilde{D}_{k+1} + \frac{\epsilon}{\tau^{k-1}}.$$

В силу $\tau > 1$ последовательность $\{\epsilon/\tau^{\nu-1}\}$ стремится к нулю, так что последовательности $\{\tilde{D}_\nu\}$ и $\{\bar{D}_\nu\}$ имеют одинаковые свойства сходимости.

Впрочем, из (6.45) получим

$$\tilde{D}_{\nu+1} + \tilde{D}_{\nu+2} = \frac{1}{\tau}\,\tilde{D}_\nu + \frac{1}{\tau^2}\,\tilde{D}_\nu = \left(\frac{1}{\tau} + \frac{1}{\tau^2}\right)\tilde{D}_\nu = \tilde{D}_\nu$$

$$\forall\, \nu \in \mathbf{N}, \tag{6.47}$$

что согласуется с (6.32). Поэтому приведенный на рис. 6.13 порядок действий может быть сохранен, если вместо D_ν применить $\tilde{D}_\nu$ и отказаться от окончания процесса при $\nu = k$.

Ввиду (6.45) и (6.47) $\tilde{D}_{\nu+1}$, $\tilde{D}_{\nu+2}$ представляют собой длины отрезков, получаемых из интервала длины $\tilde{D}_\nu$ с помощью золотого сечения. В соответствии с этим мы введем определение: пара (бесконечных) последовательностей $(\{a_\nu\}, \{b_\nu\})$ называется *парой последовательностей золотого сечения* (*GS*-парой), если при $\tilde{D}_1 = b - a$ и $\tilde{D}_\nu$, полученных при $\nu > 1$ согласно (6.45), выполняются соотношения

1) $a_1 = a$, $b_1 = b$, $\xi^{11} = b_1 - \tilde{D}_2$, $\xi^{12} = a_1 + \tilde{D}_2$;

2) $$a_{\nu+1} = \begin{cases} a_\nu & \text{при } f(\xi^{\nu 1}) < f(\xi^{\nu 2}), \\ \xi^{\nu 1} & \text{при } f(\xi^{\nu 1}) \geqslant f(\xi^{\nu 2}), \end{cases} \qquad \nu \in \mathbf{N};$$

3) $$b_{\nu+1} = \begin{cases} \xi^{\nu 2} & \text{при } f(\xi^{\nu 1}) < f(\xi^{\nu 2}), \\ b_\nu & \text{при } f(\xi^{\nu 1}) \geqslant f(\xi^{\nu 2}), \end{cases} \qquad \nu \in \mathbf{N};$$

4) $$\xi^{\nu+1,\,l} = \begin{cases} b_{\nu+1} - \tilde{D}_{\nu+2} & \text{при } a_\nu = a_{\nu+1}, \\ a_{\nu+1} + \tilde{D}_{\nu+2} & \text{при } b_\nu = b_{\nu+1}, \end{cases} \qquad \nu \in \mathbf{N},\ l = 1, 2,$$

и $\xi^{\nu+1,\,l} \neq \xi^{\nu j}$, где $j \in \{1, 2\}$ – наименьший номер, для которого $f(\xi^{\nu j}) = \min\{f(\xi^{\nu 1}), f(\xi^{\nu 2})\}$.

Таблица 6.3

ν	$\widetilde{D}_{\nu+1} = \frac{1}{\tau}\widetilde{D}_\nu$	a_ν	$\xi^{\nu 1} = \begin{cases} b_\nu - \widetilde{D}_{\nu+1} \\ \xi^{\nu-1,2} \end{cases}$	$f(\xi^{\nu 1})$	$\xi^{\nu 2} = \begin{cases} \xi^{\nu-1,1} \\ a_\nu + \widetilde{D}_{\nu+1} \end{cases}$	$f(\xi^{\nu 2})$	b_ν	$\widetilde{D}_\nu = b_\nu - a_\nu$
1	1,233	–0,50	1,50 – 1,233 = 0,267	0,073	–0,50 + 1,233 = 0,733	0,53	1,50	2,00
2	0,762	–0,50	0,733 – 0,762 = –0,029	0,0009	0,267	0,073	0,733	1,233
3	0,470	–0,50	0,267 – 0,470 = –0,203	0,040	–0,029	0,0009	0,267	0,767
4	0,290	–0,203	–0,029	0,0009	–0,203 + 0,290 = 0,087	0,008	0,267	0,470
5	0,179	–0,203	0,087 – 0,179 = –0,092	0,008 $< \delta$	–0,029	0,0009	0,087	0,290
6		–0,092					0,087	0,179

Пример 6.8. Поясним порядок действий в данном методе на задаче из примера 6.7. Не задавая наперед натуральное число k, но применяя значение δ как критерий окончания счета, получим найденные в соответствии с предыдущим определением числовые значения, которые приведены в табл. 6.3. Связи между этими значениями изображены стрелками.

В результате имеем $x^0 \in [a_6, b_6] = [-0{,}092, 0{,}087]$. $f(x^0) \leqslant \min\{f(\xi^{51}), f(\xi^{52})\} = 0{,}0009$. Для наглядности можно обратиться к рис. 6.15.

В подтверждение того, что GS-пара является парой последовательностей, порожденных последовательным сравнением значений функции в двухточечных блоках, мы покажем, что

$$\tilde{D}_\nu > \tilde{D}_{\nu+1} > \frac{1}{2}\,\tilde{D}_\nu.$$

Очевидно, эти неравенства получаются из соотношений

$$\tilde{D}_\nu = \tilde{D}_{\nu+1} + \tilde{D}_{\nu+2} > \tilde{D}_{\nu+2}, \quad \tilde{D}_{\nu+2} = \frac{1}{\tau}\,\tilde{D}_{\nu+1} \approx 0{,}618\,\tilde{D}_{\nu+1}.$$

Согласно (6.34′) и неравенству $\tilde{D}_\nu > \tilde{D}_{\nu+1}$ $\forall\, \nu \in \mathbf{N}$, GS-пара $(\{a_\nu\}, \{b_\nu\})$ порождает семейство вложенных сегментов $[a_{\nu+1}, b_{\nu+1}] \subset [a_\nu, b_\nu]$ $\forall\, \nu \in \mathbf{N}$, с помощью которых определяется ровно одно число $\bar{x} \in [a_\nu, b_\nu]$ $\forall\, \nu \in \mathbf{N}$.

Таким образом, справедлива следующая теорема.

Теорема 6.7. *Пусть функция* $f\colon [a, b] \to \mathbf{R}$ *квазиунимодальна на сегменте* $[a, b]$. *Тогда последовательность вложенных интервалов, порожденная* GS*-парой* $(\{a_\nu\}, \{b_\nu\})$, *определяет решение* x^0 *задачи* (6.23).

Доказательство. GS-пара является парой последовательностей, порожденных последовательным сравнением значений функции в двухточечных блоках. По теореме 6.4 для квазиунимодальной целевой функции существует решение $x^0 \in [a_\nu, b_\nu]$ $\forall\, \nu \in \mathbf{N}$. Последовательность вложенных интервалов определяет ровно одно число $\bar{x} \in [a_\nu, b_\nu]$ $\forall\, \nu \in \mathbf{N}$. Таким образом, $\bar{x} = x^0$.

Для того чтобы сравнить эффективность приведенных выше методов последовательного сравнения значений функции, мы рассмотрим частное длин интервалов $\tilde{D}_k$ и D_k после k шагов. Согласно (6.34) для $\nu = k$ имеет место равенство

$$D_k = (b - a)\left(\frac{1}{F_{k+1}} + \frac{F_{k-1}}{F_{k+1}}\,\delta\right).$$

Далее,

$$\tilde{D}_k = \frac{1}{\tau}\,\tilde{D}_{k-1} = \left(\frac{1}{\tau}\right)^{k-1}(b - a).$$

Считая, что D_k очень велико по сравнению с $\dfrac{F_{k-1}}{F_{k+1}}\,\delta$, имеем

$$\frac{\tilde{D}_k}{D_k} \approx \frac{F_{k+1}}{\tau^{k-1}} \approx \frac{1}{5}\left(\frac{\sqrt{5}+1}{2}\right)^2 \approx 1{,}1708.$$

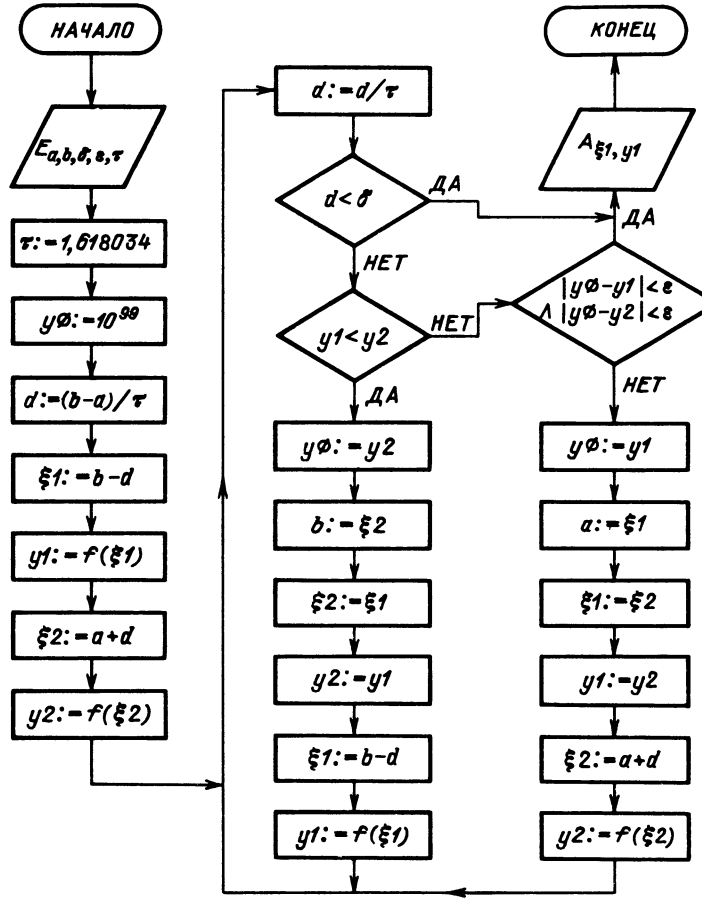


Рис. 6.16.

Несколько меньшей эффективности при применении GS-пар противостоит то преимущество, что не надо до начала вычислений оценивать число итерационных шагов.

При вычислительной реализации следует учитывать, что условие $\tilde{D}_k < \delta$ ведет к остановке метода (см. пример 6.8) и значения a_ν, b_ν, $\xi^{\nu 1}$, $\xi^{\nu 2}$ должны храниться до тех пор, пока они необходимы для последующих итераций. Этим объясняется упразднение некоторых номеров в ходе работы программы метода золотого сечения (рис. 6.16).

Так как требуется найти лишь какое-нибудь решение x^0, то в случае квазиунимодальной (но не унимодальной) целевой функции метод может остановиться, прежде чем выполнится условие $\tilde{D}_\nu < \delta$.

Теорема 6.8. *Пусть функция* $f\colon [a, b] \to \mathbf{R}$ *квазиунимодальна на множестве* $[a, b]$. *Если числа* ξ^1, ξ^2, ξ^3 *таковы, что* $a \leqslant \xi^1 < \xi^2 < \xi^3 \leqslant b$, $f(\xi^1) = f(\xi^2) = f(\xi^3)$, *то* $\xi^1, \xi^2, \xi^3 \in I^*$, *где* I^* *– множество решений задачи* (6.23).

Доказательство. Пусть $\xi^i \notin I^*$ для некоторого $i \in \{1, 2, 3\}$. Тогда из условий теоремы следует $\xi^1, \xi^2, \xi^3 \notin I^*$. Последнее же противоречит квазиунимодальности f на $[a, b]$, в силу которой функция f строго монотонна вне множества I^*.

При численной реализации предыдущее утверждение предполагает проверку условий

$$|f(\xi^1) - f(\xi^2)| < \epsilon, \quad |f(\xi^1) - f(\xi^3)| < \epsilon,$$

где ϵ есть расстояние между соседними машинными числами в представлении значений $f(\xi^i)$.

6.4. Методы решения задач с ограничениями

Вначале мы дадим обзор методов решения общей задачи нелинейного программирования

$$P_1: \min\{f(x): \ x \in G\}, \qquad G = \{x \in \mathbf{R}^n: \ g_i(x) \leqslant 0, \quad i = 1, \ldots, m\}.$$

Что касается методов решения конкретных классов нелинейных задач (например, задач гиперболического, квадратичного, сепарабельного программирования), то мы отсылаем интересующихся к работам [37, 49, 61].

Как уже пояснялось в § 6.1, метод решения задачи P_1 порождает, вообще говоря, бесконечную последовательность $\{x^k\}$ такую, что $x^k \to x^0 \in G^0$ при $k \to \infty$. Такие методы можно разбить на следующие классы.

1. Методы, которые непосредственно применяются к задаче P_1. Сюда мы отнесем, например, такие методы, на которые можно смотреть как на естественные обобщения изложенных в § 6.2 градиентных методов. Итерационные точки x^k при этом все являются допустимыми в задаче P_1.

2. Методы, которые основаны на решении последовательности вспомогательных задач H_k, при этом решения x^k задач H_k сходятся при $k \to \infty$ к $x^0 \in G^0$.

Вспомогательные задачи H_k могут быть, к примеру, задачами линейного программирования, которые возникают путем замены целевой функции и функций ограничений задачи P_1 на линейные или кусочно-линейные функции.

Другая возможность заключается в построении последовательности оптимизационных задач H_k без ограничений, к которым можно применить методы, рассмотренные в § 6.2.

Обратимся к первому классу методов. Рекуррентное соотношение

$$x^{k+1} = x^k + \alpha_k z^k, \quad \alpha_k > 0 \quad \forall k \geqslant 1, \tag{6.48}$$

которое при $z^k = -\nabla f(x^k)$ лежит в основе градиентных методов, не будет пригодным для определения решения задачи P_1, поскольку для граничной точки x^k множества G, не являющейся решением P_1, и для $z^k = -\nabla f(x^k)$, вообще говоря, не существует числа $\alpha_k > 0$ такого, что $x^{k+1} \in G$. Поэтому соотношение (6.48) естественными (но различными) способами модифицируется в методах проекции и методах возможных направлений.

Методы проекции основываются на соотношении

$$x^{k+1} = \Pi_G(x^k - \alpha_k \nabla f(x^k)),$$

где $\Pi_G(x)$ означает проекцию вектора x на множество G. При определении проекции требуется, вообще говоря, минимизировать квадратичную функцию на множестве G. Если G имеет специальную структуру, то этот подход может привести к численно эффективным алгоритмам ([98]). Другие варианты методов проекции приведены в работе [97]. Левитин и Поляк показали [98], что методы проекции сходятся со скоростью геометрической прогрессии.

Методы возможных направлений базируются на следующем соображении: при движении из точки x^k согласно (6.48) вектор z^k и число α_k выбираются таким образом, чтобы точка x^{k+1} была допустимой и $f(x^{k+1}) < f(x^k)$. Будем говорить, что вектор z^k является возможным направлением в точке $x^k \in G$, если для x^k существует такое число $\bar{\alpha} > 0$, что $x^k + \alpha z^k \in G$ для всех $\alpha \in [0, \bar{\alpha}]$. Возможное в точке x^k направление называется подходящим, если найдется число $\bar{\bar{\alpha}} > 0$ такое, что $f(x^k + \alpha z^k) < f(x^k)$ для всех $\alpha \in (0, \bar{\bar{\alpha}}]$.

Имеется большое число методов возможных направлений; их теория была обоснована Зойтендейком [87]. Вначале мы обсудим один вариант этих методов, предложенный Зойтендейком. Для облегчения понимания мы предположим, что все функции ограничений задачи P_1 являются аффинными:

$$g_i(x) = a_i' x - b_i, \quad i = 1, \ldots, m. \tag{6.49}$$

Пусть I_0^k – множество индексов активных в точке x^k ограничений. Согласно главе 4 (см. (4.26)) условия Куна – Таккера выполняются в точке x^k тогда и только тогда, когда задача линейного программирования

$$\min \{ \nabla f(x^k)' z: \quad a_i' z \leqslant 0, \quad i \in I_0^k \} \tag{6.50}$$

имеет решение $z^k = 0$. Легко догадаться (см. также рис. 6.17), что вектор z, удовлетворяющий неравенствам $a_i' z \leqslant 0$ для $i \in I_0^k$, является возможным направлением в точке x^k. Если, кроме того, для такого z выполняется $\nabla f(x^k)' z < 0$, то (см. также § 6.1) для достаточно малого α справедливо соотношение $f(x^k + \alpha z) < f(x^k)$. Следовательно, вектор z будет подходящим направлением в точке x^k. Если в (6.50) в качестве дополнительного ограничения ввести условие нормировки для векторов z (например, $\| z \| \leqslant 1$), то в результате решения возникшей задачи

$$\min \{ \nabla f(x^k)' z: \quad a_i' z \leqslant 0, \quad i \in I_0^k, \quad \| z \| \leqslant 1 \}$$

мы тогда и только тогда получим подходящее направление z^k, когда в точке x^k не будут выполняться условия Куна – Таккера.

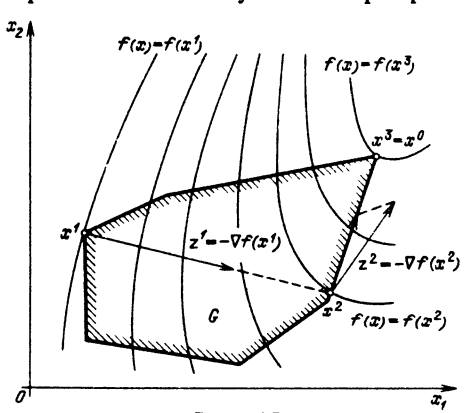


Рис. 6.17.

При соответствующем выборе α_k получаем $x^{k+1} \in G$, и, таким образом, начиная с $x^1 \in G$, мы определим последовательность допустимых точек x^k. В § 6.5 эта

основная идея переносится на случай нелинейных ограничений, и приводится доказательство сходимости.

Пиронно и Полак [68] доказали для конкретных методов возможных направлений линейную скорость сходимости.

В качестве недостатка метода возможных направлений Зойтендейка можно отметить то обстоятельство, что на каждом шаге надо решать задачу линейного программирования. Этого недостатка лишены методы проектируемых и приведенных градиентов.

Метод проектируемых градиентов Розена для задач с линейными ограничениями (который не включается в упомянутый выше класс проекционных методов) отыскивает точку z^k в случае $I_0^k \neq \phi$ как проекцию вектора $-\nabla f(x^k)$ на подпространство, заданное пересечением гиперплоскостей активных ограничений. Подробное изложение метода можно найти в работах [37, 49, 56, 69]. Кроме того, в [56] рассмотрены вопросы сходимости метода.

К хорошо зарекомендовавшим себя методам возможных направлений для задач с линейными ограничениями относится *метод приведенных градиентов* [56, 75], в котором переменные задачи подразделяются на базисные и небазисные аналогично тому, как это принято в линейном программировании. На каждом шаге метода градиент целевой функции рассматривается как функция от небазисных переменных ("приведенный градиент"), и подходящее направление находится с учетом знаков у небазисных переменных.

Фокке [30] показал, что метод приведенных градиентов при соответствующем выборе нормы z в задаче (6.50) выводится из метода Зойтендейка (см. в связи с этим § 6.5). Рассмотрение скорости сходимости метода можно найти в работе [56].

Описанные выше методы возможных направлений используют производные целевой функции и функций ограничений лишь первого порядка. В целях достижения более высокой скорости сходимости разработаны также методы возможных направлений, применяющие производные второго порядка ([69, 101]).

Теперь мы отметим некоторые методы из второго класса, введенного в начале параграфа.

В основе *метода штрафов* ([7, 27, 40, 54, 56, 69, 86, 97, 101]) лежит идея введения "штрафа" за нарушение ограничений задачи P_1. Для этого вместо P_1 решают последовательность задач без ограничений

$$H_k\colon \min\{f(x) + r_k S(x)\colon \quad x \in \mathbf{R}^n\},$$

где

$$S(x) = \sum_{i=1}^{m} g_i^{+2}(x), \qquad g_i^+(x) = \max\{0, g(x)\}.$$

Очевидно, $S(x) = 0$ тогда и только тогда, когда $x \in G$.

При соответствующих условиях решения x^k задач H_k сходятся при $r_k \to \infty$ к некоторому $x^0 \in G^0$.

На рис. 6.18 для задачи P_1, у которой $f(x) = -e^x, g_1(x) = 1 - x, g_2(x) = x^2 - 4$, $x \in \mathbf{R}$, изображена целевая функция $F(x; r_k) = -e^x + r_k (\max\{0, 1 - x\})^2 + r_k (\max\{0, x^2 - 4\})^2$, $x \in \mathbf{R}$, вспомогательной задачи при $r_k = 1, 10, 100$.

Итерационные точки x^k не являются не только оптимальными, но и допустимыми в задаче P_1.

Метод штрафов может рассматриваться как "двойственный метод" для решения P_1. Действительно, если записать задачу P_1 в эквивалентном виде

$$\tilde{P}_1 : \min \{f(x):\ S(x) \leqslant 0\},$$

то двойственной к $\tilde{P}_1$ задачей является

$$\tilde{D}_1 : \max \{\varphi(r):\ r \geqslant 0\}, \quad \varphi(r) = \inf_x \{f(x) + rS(x)\}.$$

В то время как в методе штрафов каждая не оптимальная в задаче P_1 итерационная точка не является также и допустимой для P_1, то, напротив, в *методе барьеров* ([6, 27, 56, 69, 86, 101]) все итерационные точки допустимы в P_1. Вспомогательные задачи H_k при этом конструируются так, чтобы целевые функции задач P_1 и H_k не отличались значительно внутри допустимой области G, но при приближении к границе G происходило сильное изменение целевой функции задачи H_k. Такого рода "барьеры" обеспечивают принадлежность решений вспомогательных задач H_k внутренности допустимой области, и с такими задачами можно обращаться как с задачами без ограничений.

В качестве примера рассмотрим вспомогательные задачи

$$H_k : \min \{f(x) + q_k B(x):\ g_i(x) < 0, \quad i = 1, \ldots, m\},$$

при этом $B(x) = -\sum_{i=1}^{m} \ln(-g_i(x))$ и для задачи P_1 должно быть выполнено условие $\operatorname{int} G = \{x:\ g_i(x) < 0, \quad i = 1, \ldots, m\} \neq \phi$. Очевидно, что $B(x) \to +\infty$, если $g_i(x) \to -0$ для некоторого $i \in \{1, \ldots, m\}$.

На рис. 6.19 для задачи P_1 с $f(x) = -e^x$, $g_1(x) = 1 - x$, $g_2(x) = x^2 - 4$, $x \in \mathbf{R}$, изображена целевая функция $F(x; q_k) = -e^x - q_k \ln(x - 1) - q_k \ln(4 - x^2)$, $x \in \mathbf{R}$, вспомогательной задачи при $q_k = 10$, 1 и 0,1.

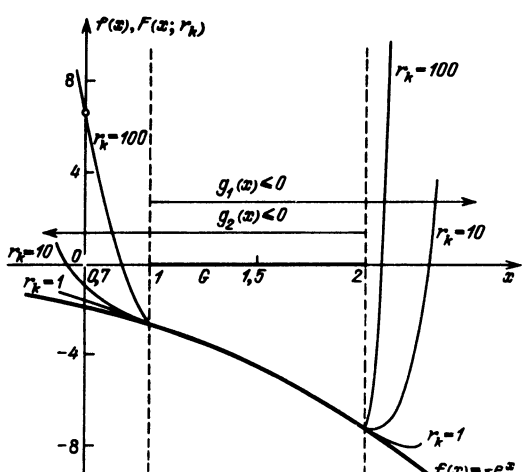


Рис. 6.18.

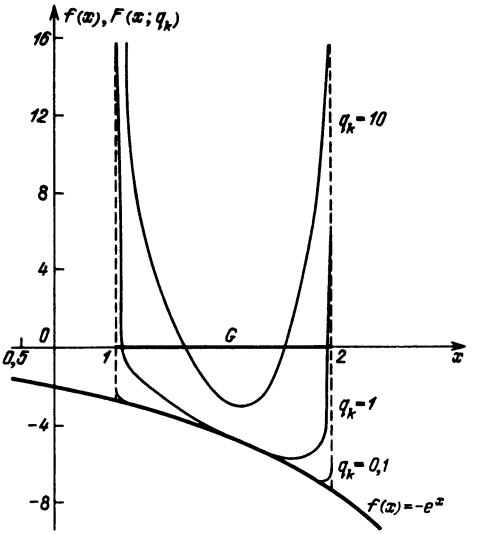


Рис. 6.19.

В § 6.6 для общего метода барьеров мы исследуем сходимость последовательности $\{x^k\}$ к решению P_1 при $q_k \to 0$.

Вычислительные затруднения при использовании метода штрафов привели к появлению метода модифицированных функций Лагранжа, который, как и метод штрафов, может характеризоваться как двойственный.

В качестве представителя большого числа работ, посвященных этому новому эффективному методу, следует назвать статью Гольштейна и Третьякова [91].

Методы секущих плоскостей основываются на решении последовательности задач линейного программирования и обычно применяются к задачам вида

$$\min\{c'x:\ g_i(x) \leqslant 0,\quad i = 1, \ldots, m\}. \tag{6.51}$$

Линейность целевой функции не ограничивает общности рассмотрения. Действительно, если требуется решить задачу вида P_1, то, добавляя ограничение $f(x) - x_{n+1} \leqslant 0$ и минимизируя (линейную) целевую функцию x_{n+1}, мы получим задачу в форме (6.51), правда, с $n+1$ переменными и $m+1$ ограничениями. Если точка $x^0 \in \mathbf{R}^n$ является решением задачи P_1, то (x^0, x^0_{n+1}), где $x^0_{n+1} = f(x^0)$, будет решением задачи (6.51) и наоборот.

Метод секущих плоскостей Келли [44] вырабатывает, начиная с некоторого ограниченного многогранного множества (многогранника) $M_1 \supseteq G$, последовательность многогранников $M_k \subseteq \mathbf{R}^n$ и последовательность точек $\{x^k\}$ по правилу, состоящему из следующих шагов.

Шаг 0. $M_1 \supseteq G$ задан. Положить $k := 1$.

Шаг 1. Вычислить решение x^k (линейной) задачи

$$P_L^k: \min\{c'x:\ x \in M_k\}.$$

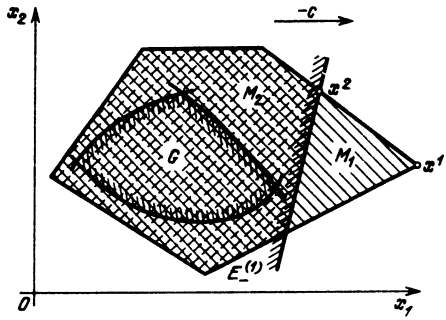


Рис. 6.20.

Ш а г 2. Если $x^k \in G$, то остановиться. В противном случае положить $M_{k+1} = M_k \cap E_-^{(k)}$, где $E_-^{(k)} = \{x \in \mathbf{R}^n: g_{i_0}(x^k) + (x - x^k)'\nabla g_{i_0}(x^k) \leqslant 0\}$ и $g_{i_0}(x^k) = \max\limits_{1 \leqslant i \leqslant m} g_i(x^k)$.

Ш а г 3. Положить $k := k + 1$ и перейти к шагу 1.

Очевидно, что поставленная в соответствие полупространству $E_-^{(k)}$ гиперплоскость

$$H^{(k)} = \{(x, \xi)' \in \mathbf{R}^{n+1}: \quad \xi = g_{i_0}(x^k) + (x - x^k)'\nabla g_{i_0}(x^k)\}$$

является одновременно касательной гиперплоскостью к гиперповерхности $\xi = g_{i_0}(x)$ в точке $x = x^k$, и для выпуклых функций g_i, $i = 1, \ldots, m$, выполняется $x^k \notin E_-^{(k)}$ и $M_k \supseteq G$ для $k = 1, 2, \ldots$ (рис. 6.20). Таким образом, на каждом шаге метода от допустимой области задачи (6.51) "отсекается" некоторая часть.

В § 6.7 будет исследован общий метод секущих плоскостей.

Оценки скорости сходимости показывают, что метод секущих плоскостей сходится медленнее, чем с порядком 1 ([98]).

Если в задаче P_1 целевую функцию и функции ограничений заменить кусочно–линейными функциями, то возникающую таким образом вспомогательную задачу, как и в методах секущих плоскостей, можно решать алгоритмами линейного программирования. Но процесс кусочной линеаризации имеет практическое значение лишь для задач с сепарабельными функциями и поэтому подробно здесь не обсуждается.

6.5. Методы возможных направлений

Пусть задача минимизации имеет следующий вид:

$$P: \min\{f(x): \ x \in G\}, \quad G = \{x \in X: \ g_i(x) \leqslant 0, \ i \in I\},$$

где $I = \{1, \ldots, m\}$, $X \subseteq \mathbf{R}^n$ – открытое множество, функции f, g_i, $i \in I$, непрерывно дифференцируемы на X. Как и в гл. 4, разобьем множество индексов I на подмножества I^L (индексы линейных ограничений) и I^N (индексы нелинейных ограничений).

Для $\epsilon \geqslant 0$ и $x^k \in G$ определим множество

$$I_\epsilon(x^k) = \{i \in I: \ -\epsilon \leqslant g_i(x^k) \leqslant 0\}. \tag{6.52}$$

Если $\epsilon > 0$, то $I_\epsilon(x^k)$ называется множеством индексов ϵ-активных ограничений в точке x^k.

Далее мы изложим метод возможных направлений, разработанных Зойтендейком (см. [87, 69, 101]). В этом методе подходящее направление z^k в точке x^k при заданном ϵ находится из решения следующей задачи:

$$P^1_{\epsilon,k}: \min\{\sigma:\ \nabla f(x^k)'z \leqslant \sigma,\ \nabla g_i(x^k)'z \leqslant \sigma,\quad i \in I^N_\epsilon(x^k);$$

$$\nabla g_i(x^k)'z \leqslant 0,\quad i \in I^L_\epsilon(x^k),\quad z \in C\}. \tag{6.53}$$

При этом C – некоторая компактная окрестность нуля. В частности, можно выбрать $C = \{z \in \mathbf{R}^n:\ \|z\| \leqslant 1\}$, где $\|\cdot\|$ – произвольная норма в $\mathbf{R}^n$. При $\|z\| = \max\limits_{1 \leqslant j \leqslant n} |z_j|$ или $\|z\| = \sum\limits_{j=1}^{n} |z_j|$ $P^1_{\epsilon,k}$ будет задачей линейного программирования. При $\epsilon = 0$, $I^N = \phi$ и $C = \{z:\ |z_j| \leqslant 1,\ j = 1, \ldots, n\}$ задача $P^1_{\epsilon,k}$ соответствует приведенной в § 6.4 задаче отыскания подходящего направления. Решение задачи $P^1_{\epsilon,k}$ обозначим через $\sigma_k = \sigma(x^k, \epsilon)$ и z^k. Имеет смысл записать задачу $P^1_{\epsilon,k}$ в следующем эквивалентном виде:

$$P^1_{\epsilon,k}:\ \min\{\max\{\nabla f(x^k)'z,\ \max_{i \in I^N_\epsilon(x^k)}\{\nabla g_i(x^k)'z\}\}:\ z \in C \cap C_\epsilon\}, \tag{6.54}$$

где

$$C_\epsilon = \{z \in \mathbf{R}^n:\ \nabla g_i(x^k)'z \leqslant 0,\quad i \in I^L_\epsilon(x^k)\}. \tag{6.55}$$

Т е о р е м а 6.9. *Для* $x^k \in G$ *и* $\epsilon = \epsilon_k \geqslant 0$ *имеют место утверждения*

1) $\sigma_k = \sigma(x^k, \epsilon_k) \leqslant 0$;

2) $\sigma(x^k, 0) \leqslant \sigma(x^k, \epsilon_k)$.

Д о к а з а т е л ь с т в о 1. Первое утверждение выполняется, поскольку $\sigma = 0$, $z = 0$ всегда будет допустимой точкой для $P^1_{\epsilon,k}$.

2. Очевидно, $I_0(x^k) \subset I_\epsilon(x^k)$ и тем самым $C_\epsilon \subset C_0$. Таким образом, $C \cap C_\epsilon \subseteq C \cap C_0$ и в силу (6.54) справедливо второе утверждение.

Используя задачу $P^1_{\epsilon,k}$, мы сформулируем как обобщение сказанного в § 6.4 критерий оптимальности, который адаптирован к методу возможных направлений.

Т е о р е м а 6.10. *Пусть в точке* x^k *выполнено условие регулярности* R_3: *существует* $\tilde{z} \in \mathbf{R}^n$ *такое, что*

$$\nabla g_i(x^k)'\tilde{z} < 0,\quad i \in I^N_0;\quad \nabla g_i(x^k)'\tilde{z} \leqslant 0,\quad i \in I^L_0. \tag{6.56}$$

Тогда:

1) *условия Куна – Таккера* (4.20), (4.21) *тогда и только тогда выполняются в точке* x^k, *когда* $\sigma(x^k, 0) = 0$;

2) *если для* $\epsilon \geqslant 0$ *точка* x^k *удовлетворяет неравенству* $\sigma(x^k, \epsilon) < 0$, *то* z^k *– подходящее направление.*

Д о к а з а т е л ь с т в о 1. Покажем, что неравенство $\sigma(x^k, 0) < 0$ имеет место тогда и только тогда, когда условия Куна – Таккера (4.20), (4.21) не выполняются в точке x^k. Соотношение $\sigma(x^k, 0) < 0$ означает,

что система

$$\begin{aligned} &\nabla f(x^k)'z < 0, \\ &\nabla g_i(x^k)'z < 0, \quad i \in I_0^N(x^k), \\ &\nabla g_i(x^k)'z \leqslant 0, \quad i \in I_0^L(x^k), \end{aligned} \tag{6.57}$$

имеет решение. Если выполнено условие R_3, то (6.57) имеет решение тогда и только тогда, когда система

$$\begin{aligned} &\nabla f(x^k)'z < 0, \\ &\nabla g_i(x^k)'z \leqslant 0, \quad i \in I_0(x^k), \end{aligned} \tag{6.58}$$

разрешима, что мы сейчас и покажем.

Пусть $\tilde{z}$ – решение (6.56), а $\bar{\bar{z}}$ – решение (6.58). Тогда $\bar{z} = \tilde{z} + \mu\bar{\bar{z}}$, где $\mu > \max\left\{0, \dfrac{\nabla f(x^k)'\tilde{z}}{-\nabla f(x^k)'\bar{\bar{z}}}\right\}$, будет решением (6.57). Обратно, любое решение системы (6.57) будет, очевидно, и решением (6.58). По теореме Фаркаша (теорема 2.43) система (6.58) разрешима тогда и только тогда, когда система

$$\begin{aligned} &\nabla f(x^k) + \sum_{i \in I_0(x^k)} \nabla g_i(x^k) y_i = 0, \\ &y_i \geqslant 0, \quad i \in I_0(x^k), \end{aligned} \tag{6.59}$$

не имеет решения, а эта система эквивалентна системе условий Куна – Таккера (4.20), (4.21).

2. Из $\sigma(x^k, \epsilon) < 0$ по теореме 6.9 получаем $\sigma(x^k, 0) < 0$. Поэтому существует решение z^k системы (6.57). Как и в доказательстве теоремы 4.7, можно отыскать число $\bar{\mu} > 0$ такое, что $x^k + \bar{\mu}z^k \in G$. В силу $\nabla f(x^k)'z^k < 0$ найдется тогда число $\bar{\bar{\mu}}$, для которого $f(x^k + \mu z^k) < f(x^k)$ при $0 < \mu \leqslant \bar{\bar{\mu}} \leqslant \bar{\mu}$. Следовательно, z^k – подходящее направление.

Сформулируем следующий алгоритм метода возможных направлений.

Шаг 0. Выбрать $\epsilon_1 > 0$ и (достаточно большое) $\eta > 0$.

Шаг 1. Определить $x^1 \in G$. Положить $k := 1$.

Шаг 2. Вычислить $\sigma_k = \sigma(x^k, \epsilon_k)$ и z^k как решение задачи $P^1_{\epsilon, k}$.

Шаг 3. Если $\sigma(x^k, \epsilon_k) = 0$, то перейти к шагу 5, иначе перейти к шагу 4.

Шаг 4. Положить

$$\epsilon_{k+1} = \begin{cases} \epsilon_k, & \text{если } \sigma_k \leqslant -\epsilon_k, \\ \epsilon_k/2, & \text{если } \sigma_k > -\epsilon_k, \end{cases}$$

и перейти к шагу 6.

Шаг 5. Если $\sigma(x^k, 0) = 0$, то положить $\hat{x} = x^k$ и остановиться. В противном случае положить $\epsilon_k = \epsilon_k : 2$ и перейти к шагу 2.

Шаг 6. Вычислить $\mu_k = \max\{\mu : x^k + \alpha z^k \in G, \ \alpha \in [0, \mu]; \ \mu \leqslant \eta\}$.

Шаг 7. Вычислить число λ_k, удовлетворяющее условию

$$f(x^k + \lambda_k z^k) = \min_{\lambda \in [0, \mu_k]} f(x^k + \lambda z^k).$$

Шаг 8. Положить $x^{k+1} = x^k + \lambda_k z^k$, $k := k + 1$ и перейти к шагу 2.

Замечания 1. Теорема 6.10 обосновывает следующее существенное свойство методов возможных направлений. Если P не является задачей

выпуклого программирования, то последовательность итерационных точек сходится, вообще говоря, лишь к точке, удовлетворяющей условиям Куна – Таккера (4.20), (4.21).

2. Для обеспечения сходимости метода при нахождении подходящего направления вместо активных ограничений в задаче $P^1_{\epsilon,k}$, должны учитываться ϵ-активные ограничения (см. также [56, 69, 101]). Но так как $\sigma(x^k, \epsilon) = 0$ согласно теоремам 6.9 и 6.10 не является достаточным требованием для выполнения условий Куна – Таккера (4.20), (4.21) в точке x^k, то для этого случая должна быть добавлена проверка соотношения $\sigma(x^k, 0) = 0$.

Обратимся теперь к исследованию сходимости метода, при этом будем следовать, в основном, работе [47].

Прежде всего выделим два случая.

Случай 1. Алгоритм после конечного числа шагов останавливается на шаге 5. Тогда по теореме 6.6, если имеет место условие регулярности R_3, то в точке $\hat{x}$ выполнены условия Куна – Таккера.

Случай 2. Алгоритм вырабатывает бесконечные последовательности $\{x^k\}$, $\{z^k\}$, $\{\sigma_k\}$, $\{\epsilon_k\}$.

Остановимся подробнее на этом случае.

Для того чтобы гарантировать существование предельной точки последовательности $\{x^k\}$, потребуем выполнение условия V: существует компактное множество M такое, что $\{x^k\} \subseteq M$.

Согласно шагу 7 алгоритма для $k = 1, 2, \ldots$ справедливо $f(x^{k+1}) < f(x^k)$. В силу условия V найдется бесконечное множество номеров K', для которого $\lim\limits_{\substack{k' \to \infty \\ k' \in K'}} x^{k'} = \hat{x}$. Тогда из свойств непрерывности функции f и монотонности последовательности $\{f(x^k)\}$ получаем

$$\lim_{\substack{k' \to \infty \\ k' \in K'}} f(x^{k'}) = \lim_{k \to \infty} f(x^k) = f(\hat{x}). \tag{6.60}$$

Основная часть доказательства сходимости состоит в том, чтобы показать, что на шаге 4 случай $-\epsilon_k < \sigma_k \leqslant 0$ встречается бесконечное число раз (т.е. для некоторого бесконечного множества номеров K). Действительно, тогда $\lim\limits_{\substack{k \to \infty \\ k \in K}} \epsilon_k = 0$, а тем самым и $\lim\limits_{\substack{k \to \infty \\ k \in K}} \sigma_k = 0$.

Теорема 6.11. *Пусть $\{x^k\}$, $\{z^k\}$, $\{\sigma_k\}$, $\{\epsilon_k\}$ – последовательности, порожденные алгоритмом возможных направлений. Тогда*

$$\lim_{k \to \infty} \epsilon_k = 0, \tag{6.61}$$

и существует бесконечное множество номеров K такое, что

$$\lim_{\substack{k \to \infty \\ k \in K}} \sigma_k = 0. \tag{6.62}$$

Доказательство. Согласно шагам 4 и 5 последовательность $\{\epsilon_k\}$ монотонно убывает и ограничена снизу. Допустим, что (6.61) не выполняется. Тогда $\lim\limits_{k \to \infty} \epsilon_k = \hat{\epsilon} > 0$ и в соответствии с шагом 4 найдется

число k_0 такое, что $\sigma_k \leqslant -\epsilon_k$ при $k \geqslant k_0$ и $\epsilon_k = \hat{\epsilon}$. Пусть $\hat{x}$ – произвольная предельная точка последовательности $\{x^k\}$. В силу $z^k \in C$ и компактности C существует бесконечное множество номеров $K'' \subseteq K'$ такое, что $\lim\limits_{k'' \to \infty} z^{k''} = \hat{z}$ и $\lim\limits_{k'' \to \infty} x^{k''} = \hat{x}$, $k'' \in K''$. Предположим вначале, что $I_0(\hat{x}) \neq I$, и рассмотрим два случая.

Случай 1: $i \in I_0(\hat{x})$. Так как $\lim\limits_{k'' \to \infty} x^{k''} = \hat{x}$, а функции g_i, $i \in I$, непрерывны, то существует число $k(i, \hat{\epsilon})$ такое, что для $k'' \geqslant k(i, \hat{\epsilon})$, $k'' \in K''$, выполняется неравенство $-\hat{\epsilon} \leqslant g_i(x^{k''}) \leqslant 0$, т.е. $i \in I_{\hat{\epsilon}}(x^{k''})$. Положим $k(\hat{\epsilon}) = \max\limits_{i \in I_0(\hat{x})} k(i, \hat{\epsilon})$. Тогда для $k'' \geqslant k(\hat{\epsilon})$, $k'' \in K''$, и $i \in I_0(\hat{x})$ имеем $i \in I_{\hat{\epsilon}}(x^{k''})$, а с учетом неравенства $\sigma_{k''} \leqslant -\epsilon_{k''}$ при $k'' \geqslant k_0$ получаем $I_0(\hat{x}) \subseteq I_{\hat{\epsilon}}(x^{k''}) = I_{\epsilon_{k''}}(x^{k''})$ для $k'' \geqslant k_1 = \max\{k_0, k(\hat{\epsilon})\}$. Поскольку (σ_k, z^k) является решением задачи $P^1_{\epsilon, k}$, то для $k'' \geqslant k_1$, $k'' \in K''$, получим

$$\begin{aligned} &\nabla f(x^{k''})' z^{k''} \leqslant \sigma_{k''} \leqslant -\epsilon_{k''} = -\hat{\epsilon}, \\ &\nabla g_i(x^{k''})' z^{k''} \leqslant \sigma_{k''} \leqslant -\epsilon_{k''} = -\hat{\epsilon}, \quad i \in I_{\epsilon_{k''}}(x^{k''}) \supseteq I_0(\hat{x}). \end{aligned} \tag{6.63}$$

Отсюда следует

$$\nabla f(\hat{x})' \hat{z} \leqslant -\hat{\epsilon}, \qquad \nabla g_i(\hat{x})' \hat{z} \leqslant -\hat{\epsilon}, \quad i \in I_0(\hat{x}). \tag{6.64}$$

Для каждого $k'' \geqslant k_1$, $k'' \in K''$, определим $\bar{\alpha}_{k''}$ как максимальное число из интервала $[0, 1]$, для которого

$$\begin{aligned} &\nabla f(x^{k''} + \alpha z^{k''})' z^{k''} \leqslant -\hat{\epsilon}/2 \quad \forall \alpha \in [0, \bar{\alpha}_{k''}], \\ &\nabla g_i(x^{k''} + \alpha z^{k''})' z^{k''} \leqslant -\hat{\epsilon}/2 \quad \forall \alpha \in [0, \bar{\alpha}_{k''}], \quad i \in I_0(\hat{x}). \end{aligned} \tag{6.65}$$

В силу (6.64) и непрерывной дифференцируемости f и g_i, $i \in I$, выполняется $\bar{\alpha}_{k''} > 0$. Более того, мы покажем, что $\gamma = \inf\limits_{k'' \to \infty} \bar{\alpha}_{k''} > 0$. Допустим, $\gamma = 0$. Тогда существует бесконечное множество номеров $K''' \subset K''$ такое, что $\lim\limits_{k''' \to \infty} \bar{\alpha}_{k'''} = 0$, $k''' \in K'''$. Без ограничения общности можно считать, что $\bar{\alpha}_{k'''} < 1$ для $k''' \in K'''$. Тогда для $k''' \in K'''$ имеем

$$\max\{\nabla f(x^{k'''} + \bar{\alpha}_{k'''} z^{k'''})' z^{k'''}, \max_{i \in I_0(\hat{x})} [\nabla g_i(x^{k'''} + \bar{\alpha}_{k'''} z^{k'''})' z^{k'''}]\} = -\frac{\hat{\epsilon}}{2}.$$

Отсюда следует

$$\max\{\nabla f(\hat{x})' \hat{z}, \max_{i \in I_0(\hat{x})} [\nabla g_i(\hat{x})' \hat{z}]\} = -\frac{\hat{\epsilon}}{2} > -\hat{\epsilon},$$

что противоречит (6.64). Поэтому $\gamma > 0$ и в силу (6.65) для $k'' \geqslant k_1$, $k'' \in K''$, имеем

$$\begin{aligned} &\nabla f(x^{k''} + \alpha z^{k''})' z^{k''} \leqslant -\hat{\epsilon}/2, \\ &\nabla g_i(x^{k''} + \alpha z^{k''})' z^{k''} \leqslant -\frac{\hat{\epsilon}}{2} \qquad \forall \alpha \in [0, \gamma]. \end{aligned} \tag{6.66}$$

Применяя теорему о среднем значении и неравенства (6.66), получим соотношения

$$f(x^{k''} + \alpha z^{k''}) = f(x^{k''}) + \alpha \nabla f(x^{k''} + \theta_0 \alpha z^{k''})' z^{k''} \leqslant f(x^{k''}) - \alpha \hat{\epsilon}/2,$$
$$g_i(x^{k''} + \alpha z^{k''}) = g_i(x^{k''}) + \alpha \nabla g_i(x^{k''} + \theta_i \alpha z^{k''})' z^{k''} \leqslant g_i(x^{k''}) - \alpha \hat{\epsilon}/2,$$
$$i \in I_0(\hat{x}), \tag{6.67}$$

справедливые для всех $\alpha \in [0, \gamma]$, любого $k'' \geqslant k_1$, $k'' \in K''$, и некоторых $\theta_0, \theta_i \in (0,1)$, $i \in I_0(\hat{x})$.

Случай 2: $i \notin I_0(\hat{x})$. Для некоторых $\epsilon_i > 0$ имеем $g_i(\hat{x}) = -2\epsilon_i < 0$. Пусть $\bar{\epsilon}$ – наименьшее из чисел ϵ_i. Тогда

$$g_i(\hat{x}) \leqslant -2\bar{\epsilon} \quad \text{для всех} \quad i \notin I_0(\hat{x}). \tag{6.68}$$

Так как функции g_i, $i \in I$, непрерывны, то существует число $\delta > 0$ такое, что $g_i(\hat{x} + \alpha\hat{z}) \leqslant \bar{\epsilon}$ для всех $\alpha \in [0, \delta]$ и $i \notin I_0(\hat{x})$.

В силу соотношений $\lim\limits_{k'' \to \infty} x^{k''} = \hat{x}$ и $\lim\limits_{k'' \to \infty} z^{k''} = \hat{z}$ найдется число $k_2 = k_2(\delta)$ такое, что для любого $\alpha \in [0, \delta]$ и $k'' \geqslant k_2$, $k'' \in K''$, выполняется

$$g_i(x^{k''} + \alpha z^{k''}) \leqslant -\bar{\epsilon}/2 \quad \forall i \notin I_0(\hat{x}). \tag{6.69}$$

Из (6.67) и (6.69) следует, что для $I_0(\hat{x}) \neq I$ справедливо $x^{k''} + \beta z^{k''} \in C$ для всех $\beta \in [0, \bar{\beta}]$ и $k'' \geqslant \bar{k}$, $k'' \in K''$, где $\bar{\beta} = \min\{\gamma, \delta, \eta\}$ и $\bar{k} = \max\{k_1, k_2\}$.

В случае $I_0(\hat{x}) = I$ соотношения (6.67) выполняются для $\bar{k} = k_1$ и $\bar{\beta} = \min\{\gamma, \eta\}$.

В соответствии с шагом 6 $\mu_{k''} \geqslant \bar{\beta} > 0$ для $k'' \geqslant \bar{k}$, $k'' \in K''$, и с учетом шага 7 и (6.67) получаем

$$f(x^{k''+1}) \leqslant f(x^{k''} + \beta z^{k''}) \leqslant f(x^{k''}) - \beta\hat{\epsilon}/2 \tag{6.70}$$

для $k'' \geqslant \bar{k}$, $k'' \in K''$.

Последнее же противоречит сходимости последовательности $\{f(x^{k''})\}$, которая гарантируется соотношением (6.60). Таким образом, исходное допущение неверно и выполняется (6.61). Тогда согласно шагу 4 существует бесконечное множество индексов K такое, что $0 \geqslant \sigma_k > -\epsilon_k$ для всех $k \in K$. Отсюда в силу равенства $\lim\limits_{\substack{k \to \infty \\ k \in K}} \epsilon_k = 0$ справедливо (6.62).

Целям подготовки доказательства сходимости служит следующая теорема.

Теорема 6.12. *Пусть K' – бесконечное множество номеров такое, что*

$$\lim_{\substack{k' \to \infty \\ k' \in K'}} x^{k'} = \hat{x} \in G, \quad \sigma(\hat{x}, 0) < 0.$$

Тогда существуют числа $\bar{\epsilon} > 0$ и $k_3 = k_3(\bar{\epsilon})$, для которых

$$\sigma(x^{k'}, \epsilon) \leqslant \frac{1}{2}\sigma(\hat{x}, 0) < 0 \tag{6.71}$$

при $\epsilon \in [0, \bar{\epsilon}]$, $k' \geqslant k_3$, $k' \in K'$.

Доказательство. 1. Покажем вначале, что

а) для любого $\epsilon \geqslant 0$ существует число $k_4(\epsilon)$ такое, что $I_\epsilon(x^{k'}) \subseteq I_0(\hat{x})$ для $k' \geqslant k_4(\epsilon)$, $k' \in K'$;

б) существует число $\bar{\epsilon}$, при котором $I_\epsilon(x^{k'}) \subseteq I_0(\hat{x})$ для $\epsilon \in [0, \bar{\epsilon}]$, $k' \geqslant k_4(\epsilon)$, $k' \in K'$.

Пусть $\epsilon \geqslant 0$. Так как в случае $I_\epsilon(\hat{x}) = I$ утверждение а) очевидно, то будем считать, что $I \neq I_\epsilon(\hat{x})$. Для $i \notin I_\epsilon(\hat{x})$ выполняется $g_i(\hat{x}) < -\epsilon$ и, в силу непрерывности g_i, $i \in I$, существует число $k_4(i, \epsilon)$ такое, что $g_i(x^{k'}) < -\epsilon$ для $k' \geqslant k_4(i, \epsilon)$. Если положим $k_4(\epsilon) = \max\limits_{i \in I \setminus I_\epsilon(\hat{x})} k_4(i, \epsilon)$, то будем иметь $g_i(x^{k'}) < -\epsilon$ для всех $k' \geqslant k_4(\epsilon)$, $k' \in K'$, и $i \notin I_\epsilon(\hat{x})$. Следовательно, $i \notin I_\epsilon(x^{k'})$ для $k' \geqslant k_4(\epsilon)$, $k' \in K'$, и тем самым (а) выполнено.

Далее, из определения непосредственно вытекает $I_0(\hat{x}) \subseteq I_\epsilon(\hat{x})$. Для $i \notin I_0(\hat{x})$, $I_0(\hat{x}) \neq I$, выберем $\bar{\epsilon}$ так же, как в доказательстве теоремы 6.11, и получим $g_i(\hat{x}) = -2\bar{\epsilon}$. Поэтому в силу определения множества $I_\epsilon(\hat{x}) = \{i: -\epsilon \leqslant g_i(\hat{x})\}$ имеем $i \notin I_\epsilon(\hat{x})$ для $\epsilon \in [0, \bar{\epsilon}]$, а это означает, что

$$I_0(\hat{x}) = I_\epsilon(\hat{x}) \quad \forall \epsilon \in [0, \bar{\epsilon}]. \tag{6.72}$$

В случае $I_0(\hat{x}) = I$ соотношение (6.72) справедливо, очевидно, для произвольного $\hat{\epsilon} > 0$. Из (6.72) и а) следует б).

2. Пусть теперь $\sigma(\hat{x}, 0) < 0$. Определим функцию

$$\bar{\sigma}(x) = \min_{z \in C \cap \hat{C}_0} \max\{\nabla f(x)'z, \max_{i \in I_0^N(\hat{x})} [\nabla g_i(x)'z]\}, \tag{6.73}$$

где

$$\hat{C}_0 = \{z: \nabla g_i(x)'z \leqslant 0, \ i \in I_0^L(\hat{x})\}. \tag{6.74}$$

Для $x = \hat{x}$ имеем $\bar{\sigma}(\hat{x}) = \sigma(\hat{x}, 0) < 0$. Функция $\bar{\sigma}$ непрерывна по x. Поэтому существует число k_5 такое, что

$$\bar{\sigma}(x^{k'}) \leqslant \frac{1}{2}\bar{\sigma}(\hat{x}) = \frac{1}{2}\sigma(\hat{x}, 0) < 0 \quad \forall k' \geqslant k_5, \ k' \in K'. \tag{6.75}$$

Из б) следует $C_\epsilon \supseteq \hat{C}_0$, и согласно (6.73), (6.74) и (6.54), (6.55) получаем

$$\sigma(x^{k'}, \epsilon) \leqslant \bar{\sigma}(x^{k'}) \quad \forall \epsilon \in [0, \bar{\epsilon}], \qquad \forall k' \geqslant k_4(\epsilon), \ k' \in K'. \tag{6.76}$$

При $k_3 = \max\{k_4(\epsilon), k_5\}$ утверждение теоремы следует из (6.75) и (6.76).

Применяя предыдущие теоремы, можно доказать теперь следующее свойство сформулированного выше алгоритма.

Теорема 6.13. *Пусть последовательности* $\{x^k\}$, $\{z^k\}$, $\{\sigma_k\}$, $\{\epsilon_k\}$ *порождены алгоритмом возможных направлений и выполнены условия* V *и* R_3. *Справедливы следующие утверждения.*

1) *Существует бесконечное множество номеров* K_1, *для которого имеет место* $\lim\limits_{\substack{k \to \infty \\ k \in K_1}} x^k = \hat{x}$. *В точке* $\hat{x}$ *выполняются условия Куна – Таккера* (4.20), (4.21).

2) *Если функция* f *псевдовыпукла, а* g_i, $i \in I$, *квазивыпуклы, то любая предельная точка последовательности* $\{x^k\}$ *является решением задачи* P.

Доказательство. 1. По теореме 6.11 найдется множество номеров K такое, что $\lim\limits_{\substack{k\to\infty\\ k\in K}} \sigma_k = 0$. В силу условия V существует бесконечное множество номеров $K_1 \subseteq K$, для которого $\lim\limits_{\substack{k\to\infty\\ k\in K_1}} x^k = \hat{x}$ и $\lim\limits_{\substack{k\to\infty\\ k\in K_1}} \epsilon_k = 0$. Допустим, что в точке $\hat{x}$ не выполнены условия Куна – Таккера. Тогда по теореме 6.10 $\sigma(\hat{x}, 0) < 0$ и согласно теореме 6.12 существуют числа k_3 и $\bar{\epsilon} > 0$ такие, что

$$\sigma(x^k, \epsilon) \leqslant \frac{1}{2}\sigma(\hat{x}, 0) < 0 \qquad \forall \epsilon \in [0, \bar{\epsilon}], \quad k \geqslant k_3, \quad k \in K_1. \tag{6.77}$$

В силу (6.61) должно найтись число $k_6 \geqslant k_3$, при котором $\epsilon_k \leqslant \bar{\epsilon}$ для всех $k \geqslant k_6$, $k \in K_1$. Таким образом, из (6.77) получаем $\sigma_k = \sigma(x^k, \epsilon_k) \leqslant$ $\leqslant \frac{1}{2}\sigma(\hat{x}, 0) < 0$ для $k \geqslant k_6$, $k \in K_1$.

Последнее противоречит соотношению (6.62). Следовательно, в точке $\hat{x}$ условия Куна – Таккера (4.20), (4.21) выполняются.

2. По теореме 4.26 выполнения условий Куна – Таккера в предположениях второго утверждения доказываемой теоремы достаточно для того, чтобы $\hat{x}$ было решением задачи P. Пусть теперь $\hat{\hat{x}}$ – произвольная предельная точка последовательности $\{x^k\}$. Тогда найдется множество номеров K_2 такое, что $\lim\limits_{\substack{k\to\infty\\ k\in K_2}} x^k = \hat{\hat{x}}$. В силу (6.60)

$$f(\hat{\hat{x}}) = \lim_{\substack{k\to\infty\\ k\in K_2}} f(x^k) = \lim_{k\to\infty} f(x^k) = \lim_{\substack{k\to\infty\\ k\in K_1}} f(x^k) = f(\hat{x}).$$

Поэтому и точка $\hat{\hat{x}}$ будет решением задачи P.

Пример 6.9. Решим с помощью заданного алгоритма следующую невыпуклую задачу (рис. 6.21):

$$\min\{f(x) = (x_1 + 4)^2 + (x_2 - 7)^2 : x \in G\},$$

где $G = \{x \in \mathbb{R}^2 : g_1(x) = x_1^2 - 3{,}5x_2 \leqslant 0, \quad g_2(x) = x_1 - (x_2 - 3)^2 + 2{,}2 \leqslant 0, \quad g_3(x) = -x_1 + x_2 - 6 \leqslant 0, \; g_4(x) := x_1 - 2x_2 + 1{,}5 \leqslant 0\}$.

В качестве начальной точки выберем $x^1 = (0, 3/2)'$. Данные о ходе решения содержатся в табл. 6.4.

В точке $\hat{x} = x^4$ выполнены условия Куна – Таккера (4.20), (4.21), поскольку $\sigma(x^4, 0) = 0$. Точка $\hat{x}$ здесь является одновременно решением поставленной задачи (см. рис. 6.21).

Теперь остановимся детально на других методах возможных направлений.

Зойтендейк привел в работе [87] метод возможных направлений, который для нахождения допустимого направления использует решение z^k, $\sigma_k = \sigma(x^k, \epsilon)$ следующей задачи:

$$P^2_{\epsilon,k}: \min\{\sigma : \nabla f(x^k)'z \leqslant \sigma, \; \nabla g_i(x^k)'z + g_i(x^k) \leqslant \sigma, \; i \in I^N_\epsilon(x^k);$$

$$\nabla g_i(x^k)'z + g_i(x^k) \leqslant 0, \quad i \in I^L_\epsilon(x^k); \quad z \in C\}.$$

Таблица 6.4

Шаг 2							Шаг 3	Шаг 5	Шаг 6		Шаг 7
k	ϵ_k	x^k	$g_i(x^k)$		$\nabla g_i(x^k)$ для $i \in I_{\epsilon_k}(x^k)$	Решение $P^1_{\epsilon_k,k}$	ϵ_{k+1}	Решение $P^1_{0,k}$	μ_k	λ_k	x^{k+1}
1	2	$\begin{pmatrix}0\\1{,}5\end{pmatrix}$	$i=1$	$-5{,}25$		$z^1=\begin{pmatrix}-1\\-0{,}5\end{pmatrix}$	2	$\Longrightarrow$	1,5776771	1	$\begin{pmatrix}-1\\1\end{pmatrix}$
			$i=2$	$-0{,}05$	$(1,3)'$						
			$i=3$	$-4{,}5$							
			$i=4$	$-1{,}5$	$(1,-2)'$						
			$I_{\epsilon_1}(x^1)=\{2,4\}$		$\nabla f(x^1)=(8,-11)'$	$\sigma_1=-2{,}5$					
2	2	$\begin{pmatrix}-1\\1\end{pmatrix}$	$i=1$	$-2{,}5$		$z^2=\begin{pmatrix}-1\\1\end{pmatrix}$	2	$\Longrightarrow$	2	2	$\begin{pmatrix}-3\\3\end{pmatrix}$
			$i=2$	$-2{,}8$							
			$i=3$	-4							
			$i=4$	$-1{,}5$	$(1,-2)'$						
			$I_{\epsilon_2}(x^2)=\{4\}$		$\nabla f(x^2)=(6,-12)'$	$\sigma_2=-18$					
	2	$\begin{pmatrix}-3\\3\end{pmatrix}$	$i=1$	$-1{,}5$	$(-6,-3{,}5)'$	$z^3=\begin{pmatrix}0\\0\end{pmatrix}$	$\Longrightarrow$	$I_0(x^3)=\{3\}$			
			$i=2$	$-0{,}8$	$(1,0)'$						
			$i=3$	0	$(-1,1)'$						
			$i=4$	$-7{,}5$							
			$I_{\epsilon_3}(x^3)=\{1,2,3\}$		$\nabla f(x^3)=(2,-6)'$	$\sigma_3=0$		$\sigma(x^3,0)=-8$			

3	1	$\begin{pmatrix}-3\\3\end{pmatrix}$	$i=1$			$z^3=\begin{pmatrix}0\\0\end{pmatrix}$	$\Longrightarrow$	$I_0(x^3)=\{3\}$			
			$i=2$		$(1,0)'$						
			$i=3$		$(-1,1)'$						
			$i=4$								
			$I_{\epsilon_3}(x^3)=\{2,3\}$			$\sigma_3=0$		$\sigma(x^3,0)=-8$			
	0,5	$\begin{pmatrix}-3\\3\end{pmatrix}$	$i=1$			$z^3=\begin{pmatrix}1\\1\end{pmatrix}$	0,5	$\Longrightarrow$		2,5	$\begin{pmatrix}-1,5\\4,5\end{pmatrix}$
			$i=2$								
			$i=3$		$(-1,1)'$						
			$i=4$								
			$I_{\epsilon_3}(x^3)=\{3\}$			$\sigma_3=-8$					
4	0,5	$\begin{pmatrix}-1,5\\4,5\end{pmatrix}$	$i=1$	-10		$z^4=\begin{pmatrix}-1\\-1\end{pmatrix}$	$\Longrightarrow$	$I_0(x^4)=\{3\}$			
			$i=2$	$-0,8$							
			$i=3$	0	$(-1,1)'$						
			$i=4$	$-8,5$							
			$I_{\epsilon_4}(x^4)=\{3\}$		$\nabla f(x^4)=(5,-5)'$	$\sigma_4=0$		$\underline{\sigma(x^4,0)=0}$			

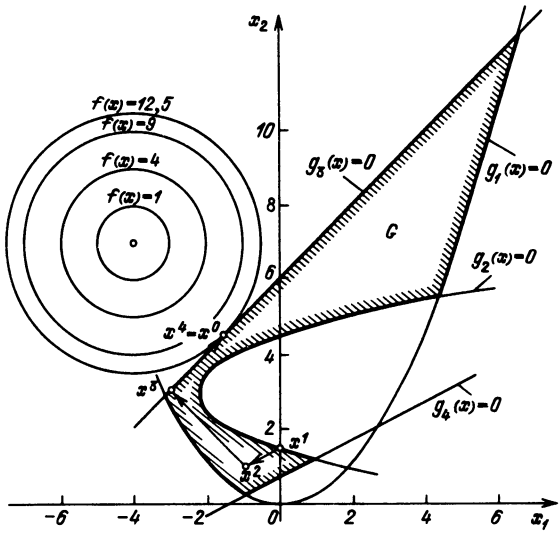


Рис. 6.21.

Соответствующий алгоритм выглядит так:

Шаг 0. Выбрать $\epsilon > 0$ и (достаточно большое) $\eta > 0$.

Шаг 1. Определить $x^1 \in G$ и положить $k = 1$.

Шаг 2. Вычислить $\sigma_k = \sigma(x^k, \epsilon)$ и z^k как решение задачи $P^2_{\epsilon,k}$.

Шаг 3. Если $\sigma_k = 0$, то положить $\hat{x} = x^k$ и остановиться. В противном случае перейти к шагу 4.

Шаг 4. Вычислить

$$\mu_k = \max\{\mu: x^k + \alpha z^k \in G,\ \alpha \in [0, \mu];\ \mu \leqslant \eta\}.$$

Шаг 5. Определить λ_k из условия

$$f(x^k + \lambda_k z^k) = \min_{\lambda \in [0, \mu_k]} f(x^k + \lambda z^k).$$

Шаг 6. Положить $x^{k+1} = x^k + \lambda_k z^k$, $k := k + 1$ и перейти к шагу 2.

В этом алгоритме $\epsilon > 0$ остается постоянным. Можно показать ([47, 69]), что при выполнении V и R_3 в любой предельной точке $\hat{x}$ последовательности $\{x^k\}$, порождаемой последним алгоритмом, справедливы условия Куна – Таккера.

Справедлива следующая теорема, аналогичная теореме 6.10

Теорема 6.10′. *Пусть в точке x^k выполнено условие регулярности R_3 и $\epsilon > 0$ произвольно. Имеют место следующие утверждения.*

1) *Условия Куна – Таккера* (4.20), (4.21) *выполняются в точке x^k тогда и только тогда, когда для задачи $P^2_{\epsilon,k}$ $\sigma_k = \sigma(x^k, \epsilon) = 0$.*

2) *Если $\sigma_k < 0$, то z^k является подходящим направлением.*

Как уже упоминалось в § 6.4, разработаны методы возможных направлений, в которых для определения подходящего направления не требуется решение оптимизационной задачи.

Метод приведенных градиентов для задачи P с $I^N = \phi$ может быть истолкован как частный случай описанного метода возможных направлений,

в котором множество C задается с помощью эллипсоидной нормы и последняя может меняться в ходе итерационного процесса.

Пусть имеется задача P с $I^N = \phi$ следующего вида:

$$\min \{ f(x): Ax - b \leqslant 0 \},$$

где A – матрица размера $m \times n$, $r(A) = n$ и функция f: $\mathbf{R}^n \to \mathbf{R}$ непрерывно дифференцируема.

Векторы-строки матрицы A обозначим через ${}^i a$, $i \in I = \{1, \ldots, m\}$. Чтобы исключить вырожденный случай, мы предположим, что векторы ${}^i a$, $i \in I_{\epsilon_k}(x^k)$, линейно независимы, так что $|I_{\epsilon_k}(x^k)| \leqslant n$. Дополним векторы ${}^i a$, $i \in I_{\epsilon_k}(x^k)$ другими $(n - |I_{\epsilon_k}(x^k)|)$ строками ${}^i a$ из A до невырожденной матрицы

$$B = \begin{pmatrix} {}^i a, & i \in I_{\epsilon_k}(x^k) \\ \ldots\ldots\ldots\ldots \\ {}^i a, & i \in I_k \end{pmatrix},$$

где $I_k \subseteq I \setminus I_{\epsilon_k}(x^k)$.

Если выбрать $C = \{ z \in \mathbf{R}^n: z' B' B z \leqslant 1 \}$, то задаче $P^1_{\epsilon,k}$ определения подходящего направления z^k будет соответствовать задача

$$\bar{P}^1_{\epsilon,k}: \min \{ \nabla f(x^k)' z: {}^i a z \leqslant 0, \ i \in I_{\epsilon_k}(x^k), \ z' B' B z \leqslant 1 \}.$$

Решение $\bar{P}^1_{\epsilon,k}$ может быть указано явно [30]. Для этого рассмотрим "приведенный" к базису B градиент функции f:

$$d = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} d^1_i, & i \in I_{\epsilon_k}(x^k) \\ \ldots\ldots\ldots\ldots \\ d^2_i, & i \in I_k \end{pmatrix} = (B^{-1})' \nabla f(x^k).$$

Полагая $w = (\| (d^1)^+ \|^2 + \| d^2 \|^2)^{1/2}$, $(d^1_i)^+ = \max \{ 0, d^1_i \}$, $i \in I_{\epsilon_k}(x^k)$, получим для $w > 0$ подходящее направление

$$z^k = -\frac{1}{w} B^{-1} \begin{pmatrix} (d^1)^+ \\ d^2 \end{pmatrix}.$$

В случае $w = 0$ в точке x^k выполняются условия Куна – Таккера.

В работе [30] предложена модификация метода приведенных градиентов, которая позволяет надеяться на более высокую эффективность. Обобщение метода на случай нелинейных ограничений можно найти в книге [51].

6.6. Методы штрафов и барьеров

6.6.1. Предварительное замечание. В этом параграфе мы рассматриваем задачу

$$P: \min \{ f(x): x \in G \subseteq \mathbf{R}^n \},$$

где f: $\mathbf{R}^n \to \mathbf{R}$ – непрерывная функция.

В основе метода штрафов (метода штрафных функций), а равно и метода барьеров (метода барьерных функций), лежит идея, состоящая

в замене исходной задачи P последовательностью $\{H_k\}$ вспомогательных задач без ограничений (см. § 6.4). При этом целевые функции вспомогательных задач выбираются таким образом, чтобы решения x^k задач H_k сходились при $k \to \infty$ к решению x^0 задачи P. Для решения вспомогательных задач привлекаются методы решения задач без ограничений (см § 6.6.2). Правда, в некоторых случаях оказывается выгодным часть ограничений, определяющих допустимое множество задачи P, сохранить в процессе образования вспомогательных задач. Это́ касается, в частности, граничных значений для переменных x_j или, вообще, линейных ограничений, которые определяют, например, множество $G_L \supseteq G$. Поэтому мы выберем следующее представление допустимой области:

$$G = G_N \cap G_L; \quad G_N, G_L \subseteq \mathbf{R}^n.$$

Тогда соответствующие вспомогательные задачи примут вид

$$H_k\colon \min \{ F_k(x)\colon x \in G_L \subseteq \mathbf{R}^n \}. \tag{6.79}$$

При этом не исключается случай $G_L = \mathbf{R}^n$.

6.6.2. Метод штрафов. В этом классе методов нарушение (части) ограничений карается "штрафами", причем эти штрафы увеличиваются с ростом номера k. Для вспомогательных задач H_k определим

$$F_k(x) = F(x; r_k) = f(x) + r_k S(x), \tag{6.80}$$

$$S\colon \mathbf{R}^n \to \mathbf{R} \text{ — непрерывная функция}, \tag{6.81}$$

$$S(x) > 0 \text{ для } x \notin G_N, \qquad S(x) = 0 \text{ для } x \in G_N. \tag{6.82}$$

От последовательности $\{r_k\}$ потребуем, чтобы

$$r_{k+1} > r_k > 0, \quad k = 1,2,\ldots, \tag{6.83}$$

$$\lim_{k \to \infty} r_k = \infty. \tag{6.84}$$

Для применения метода штрафов существенно предположение о существовании решений x^k у вспомогательных задач H_k, т.е. выполнение равенства

$$F(x^k; r_k) = \min_{x \in G_L} F(x; r_k). \tag{6.85}$$

Достаточное для этого условие содержит следующая теорема.

Т е о р е м а 6.14. *Пусть множество $G_L \subseteq \mathbf{R}^n$ замкнуто. Если существует вектор $z \in G_N$ такой, что множество*

$$\Lambda = \{ x \in G_L\colon f(x) \leqslant f(z) \} \tag{6.86}$$

компактно, то задачи H_k имеют решения x^k ($k = 1,2,\ldots$).

Д о к а з а т е л ь с т в о. В силу (6.82) и (6.83) для любого $k = 1,2,\ldots$ имеет место соотношение

$$\Lambda_k = \{ x \in \mathbf{R}^n\colon f(x) + r_k S(x) \leqslant f(z) \} \subseteq \Lambda.$$

Следовательно, Λ_k ограничено. Покажем, что Λ_k также замкнуто. С этой целью рассмотрим последовательность $\{ x^l \} \subseteq \Lambda_k$ такую, что $x^l \to y$ при $l \to \infty$. Тогда имеем $f(x^l) + r_k S(x^l) \leqslant f(z)$. Так как функции f и S непре-

рывны, то получаем

$$F(y;r_k) = \lim_{l \to \infty} F(x^l;r_k) \leqslant f(z).$$

Поскольку множество G_L замкнуто, получаем $y \in \Lambda_k$. Таким образом, Λ_k компактно и справедливо равенство

$$\inf_{x \in G_L} F(x;r_k) = \inf_{x \in \Lambda_k} F(x;r_k) = F(x^k;r_k).$$

Применяя специальный критерий окончания счета (см. теорему 6.19), мы получаем следующий алгоритм.

Шаг 0. Задать явно или с помощью рекуррентного правила последовательность $\{r_k\}$ в соответствии с (6.83), (6.84). Найти $x^1 \in G_L$. Положить $k = 1$.

Шаг 1. Положить $r = r_k$.

Шаг 2. Вычислить x^k из условия

$$F(x^k;r) = \min_{x \in G_L} F(x;r).$$

Шаг 3. Если выполнен критерий окончания счета, то остановиться. В противном случае положить $k := k + 1$ и перейти к шагу 1.

Непосредственно из структуры целевой функции $F(x;r_k)$ вспомогательной задачи H_k вытекает следующее утверждение.

Теорема 6.15. *Пусть* $\{x^k\}$ – *последовательность решений вспомогательных задач* H_k, *порождаемая методом штрафов согласно* (6.85), *и* $f(x^0)$ – *оптимальное значение задачи* P. *Справедливы соотношения*

$$1)\ F(x^k;r_k) \leqslant F(x^{k+1};r_{k+1}), \quad k = 1,2,\ldots; \tag{6.87}$$

$$2)\ f(x^k) \leqslant f(x^{k+1}), \quad k = 1,2,\ldots; \tag{6.88}$$

$$3)\ f(x^k) \leqslant F(x^k;r_k) \leqslant f(x^0), \quad k = 1,2,\ldots. \tag{6.89}$$

Доказательство. 1. Для $k = 1,2,\ldots$ имеем

$$F(x^k;r_k) = f(x^k) + r_k S(x^k) \leqslant f(x^{k+1}) + r_k S(x^{k+1}) \leqslant$$

$$\leqslant f(x^{k+1}) + r_{k+1} S(x^{k+1}) = F(x^{k+1};r_{k+1}).$$

2. Для $k = 1,2,\ldots$ справедливы неравенства

$$f(x^k) + r_k S(x^k) \leqslant f(x^{k+1}) + r_k S(x^{k+1}),$$

$$f(x^{k+1}) + r_{k+1} S(x^{k+1}) \leqslant f(x^k) + r_{k+1} S(x^k).$$

Отсюда следует

$$\left(1 - \frac{r_k}{r_{k+1}}\right) f(x^k) \leqslant \left(1 - \frac{r_k}{r_{k+1}}\right) f(x^{k+1}),$$

т.е. (6.88).

3. Из (6.82), (6.83) и (6.85) для $k = 1,2,\ldots$ имеем

$$f(x^k) \leqslant f(x^k) + r_k S(x^k) \leqslant f(x^0) + r_k S(x^0) = f(x^0).$$

Оптимальное значение $f(x^0)$ задачи P является, таким образом, верхней оценкой для неубывающих последовательностей $\{f(x^k)\}$ и $\{F(x^k;r_k)\}$.

Если итерационная точка x^k допустима в задаче P, то x^k одновременно решает и P. В этом случае метод может остановиться уже через конечное число шагов.

Теорема 6.16. *Пусть* $\{x^k\}$ – *порождаемая методом штрафов в соответствии с* (6.85) *последовательность решений задач* H_k. *Любая предельная точка последовательности* $\{x^k\}$ *есть решение задачи* P.

Доказательство. Пусть $\{x^{k_j}\}$ – сходящаяся подпоследовательность из $\{x^k\}$ и $y \in G_L$ – ее предел. Если мы покажем, что $y \in G_N \cap G_L$, то в силу (6.89)

$$f(y) = \lim_{k_j \to \infty} f(x^{k_j}) \leqslant f(x^0),$$

т.е. y – решение задачи P.

Согласно (6.87) последовательность $\{F(x^{k_j}; r_{k_j})\}$ монотонно возрастает. Так как она к тому же ограничена сверху значением $f(x^0)$, то

$$\lim_{k_j \to \infty} F(x^{k_j}; r_{k_j}) = \delta \leqslant f(x^0). \tag{6.90}$$

Из непрерывности функции f следует, что

$$\lim_{k_j \to \infty} f(x^{k_j}) = f(y). \tag{6.91}$$

В силу (6.90) и (6.91) имеем

$$\lim_{k_j \to \infty} F(x^{k_j}; r_{k_j}) - \lim_{k_j \to \infty} f(x^{k_j}) = \lim_{k_j \to \infty} r_{k_j} S(x^{k_j}) = \delta - f(y). \tag{6.92}$$

Учитывая, что $S(x^{k_j}) \geqslant 0 \quad \forall j = 1,2,\ldots$ и $r_{k_j} \to \infty$, из (6.92) получаем

$$\lim_{k_j \to \infty} S(x^{k_j}) = 0.$$

Тогда из непрерывности функции S следует $S(y) = 0$. В силу (6.82) $y \in G_N \cap G_L$.

Согласно теоремам 6.15 и 6.16 последовательности $\{f(x^k)\}$ и $\{F(x^k; r_k)\}$ монотонно сходятся к $f(x^0)$ при $k \to \infty$.

Если множество G_N задано в виде

$$G_N = \{x \in \mathbf{R}^n : g_i(x) \leqslant 0, \ i \in I\},$$

где $g_i : \mathbf{R}^n \to \mathbf{R}$ – непрерывные функции, то в качестве реализаций для функции S можно взять, например, следующие:

$$1)\ S(x) = \max_{i \in I} \{0, g_i(x)\}, \tag{6.93}$$

$$2)\ S(x) = \sum_{i=1}^{m} s(g_i(x)), \tag{6.94}$$

где

$$s(t) = 0, \quad t \leqslant 0, \qquad s(t) > 0, \quad t > 0, \tag{6.95}$$

$$s : \mathbf{R} \to \mathbf{R} \text{ – непрерывная функция.} \tag{6.96}$$

В качестве примера функции s можно привести

$$s(t) = (t^+)^p, \quad p \geqslant 1.$$

6.6.3. Метод штрафов для задач выпуклого программирования. Любая допустимая точка задачи P определяет верхнюю оценку для оптимального значения $f(x^0)$. Если задача P выпуклая, то удается простым способом последовательности $\{x^k\}$, порожденной методом штрафов, поставить в соответствие последовательность $\{z^k\}$ допустимых для P точек таких, что $\lim\limits_{k\to\infty} f(z^k) = f(x^0)$.

Предположим, что $G_N = \{x \in \mathbf{R}^n: g_i(x) \leqslant 0,\ i \in I\}$, g_i – выпуклые функции, $i \in I$ и существует точка $\bar{x} \in G_L$, удовлетворяющая неравенствам $g_i(x) < 0 \quad \forall i \in I$. Далее, положим

$$z^k = \alpha_k \bar{x} + (1 - \alpha_k)x^k, \qquad \alpha_k \in [0, 1], \tag{6.97}$$

где α_k – наименьшее из чисел α, являющихся решением неравенства

$$\alpha g_i(\bar{x}) + (1 - \alpha)g_i(x^k) \leqslant 0, \qquad i \in I,$$

т.е.

$$\alpha_k = \begin{cases} \max\limits_{i \in I_k} \dfrac{g_i(x^k)}{g_i(x^k) - g_i(\bar{x})}, \\ 0, \quad \text{если} \quad I_k = \phi, \end{cases} \tag{6.98}$$

где

$$I_k = \{i \in I: g_i(x^k) > 0\}. \tag{6.99}$$

Очевидно, что z^k является допустимой в задаче P точкой для $k = 1,2,\ldots$, поскольку $z^k \in G_L$ и в силу выпуклости g_i, $i \in I$, имеют место неравенства

$$g_i(z^k) \leqslant \alpha_k g_i(\bar{x}) + (1 - \alpha_k)g_i(x^k) \leqslant 0, \quad i \in I.$$

Пусть теперь точка y – предел сходящейся подпоследовательности $\{x^{k_j}\}$, выделенной из полученной с помощью метода штрафов последовательности $\{x^k\}$. По теореме 6.16 $y \in G_N \cap G_L$ и $f(y) = f(x^0)$. Из (6.98) в силу непрерывности функций g_i тотчас следует $\alpha_{k_j} \to 0$ при $k_j \to \infty$. Тем самым получаем $f(x^{k_j}) \to f(y)$ при $k_j \to \infty$. Следовательно, справедлива следующая теорема.

Т е о р е м а 6.17. *Пусть g_i – выпуклые функции, $i \in I$, а $\{x^k\}$ и $\{z^k\}$ – последовательности, порождаемые методом штрафов согласно* (6.97), (6.98), (6.99). *Тогда*

1) $f(x^k) \leqslant F(x^k; r_k) \leqslant f(x^0) \leqslant f(z^k), \quad k = 1,2,\ldots,$ (6.100)

2) *для любой сходящейся подпоследовательности $\{x^{k_j}\}$ из $\{x^k\}$*

$$\lim_{k_j \to \infty} f(x^{k_j}) = \lim_{k_j \to \infty} f(z^{k_j}) = f(x^0).$$

Накладывая дополнительные ограничения на условия задачи, можем охарактеризовать сходимость метода штрафов следующим образом.

Т е о р е м а 6.18. *Пусть множество $G_N \cap G_L$ компактно, G_L выпукло и замкнуто, функции g_i, $i \in I$, выпуклы, $F(x; r_k)$ выпукла по x. Предпо-*

ложим, что существует точка $\bar{x} \in G_L$, *для которой выполняется* $g_i(\bar{x}) < 0$ $\forall i \in I$. *Тогда для любого* $\delta > 0$ *найдется число* $r(\delta)$ *такое, что при* $r_k > r(\delta)$ *функция* $F(x; r_k)$ *достигает минимума на* G_L *в некоторой точке из множества* $G_N^\delta \cap G_L$, *где* $G_N^\delta = \{x \in \mathbf{R}^n: g_i(x) \leqslant \delta, i \in I\}$.

Доказательство. Пусть δ – произвольное положительное число. Функция $g: G_L \to \mathbf{R}$ вида $g(x) = \max\limits_{i \in I} g_i(x)$ выпукла и непрерывна на замкнутом множестве G_L. По условию множество $G = G_N \cap G_L = \{x \in \mathbf{R}^n: g(x) \leqslant 0\}$ компактно. Следовательно, по теореме 2.87 компактно и множество $\{x \in \mathbf{R}^n: g(x) \leqslant \delta\} = G_N^\delta \cap G_L$. Множество $\hat{G}_N^\delta = \{x \in \mathbf{R}^n: g(x) = \delta\}$ замкнуто, и в силу $\hat{G}_N^\delta \subseteq G_N^\delta$ будет компактно и множество $\hat{G}_N^\delta \cap G_L$. Отсюда имеем

$$F(x; r_k) \geqslant \min_{x \in \hat{G}_N^\delta \cap G_L} \{f(x) + r_k S(x)\} \geqslant f(x^1) + r_k S(x^2), \tag{6.101}$$

где $f(x^1) = \min\limits_{x \in \hat{G}_N^\delta \cap G_L} f(x)$, $S(x^2) = \min\limits_{x \in \hat{G}_N^\delta \cap G_L} S(x)$. Так как $\hat{G}_N^\delta \cap G_N = \phi$, то $S(x^2) > 0$, и если $r_k > r(\delta) = \dfrac{|f(\bar{x}) - f(x^1)|}{S(x^2)}$, то

$$F(x; r_k) = f(x^1) + r_k S(x^2) > f(\bar{x}) = F(\bar{x}; r_k).$$

Таким образом, при $r_k > r(\delta)$ получаем

$$F(x; r_k) > F(\bar{x}; r_k) \quad \forall x \in \hat{G}_N^\delta \cap G_L.$$

Для любой точки $\bar{\bar{x}} \in G_L \setminus G_N^\delta$ относительная внутренность отрезка $[\bar{\bar{x}}, \bar{x}]$ пересекает множество $\hat{G}_N^\delta \cap G_L$. Следовательно, для каждого $\bar{\bar{x}} \in G_L \setminus G_N^\delta$ существует число $\beta \in (0, 1)$ такое, что $x = \beta\bar{\bar{x}} + (1 - \beta)\bar{x} \in \in \hat{G}_N^\delta \cap G_L$. Так как $F(x; r_k)$ выпукла относительно x, то для $r_k > r(\delta)$ получаем

$$F(\bar{x}; r_k) < F(x; r_k) \leqslant \beta F(\bar{\bar{x}}; r_k) + (1 - \beta) F(\bar{x}; r_k) \quad \forall \bar{\bar{x}} \in G_L \setminus G_N^\delta,$$

$$F(\bar{x}; r_k) < F(\bar{\bar{x}}; r_k) \quad \forall \bar{\bar{x}} \in G \setminus G_N^\delta. \tag{6.102}$$

Из компактности множества $G_N^\delta \cap G_L$ вытекает существование точки $x^k \in G_N^\delta \cap G_L$, для которой

$$F(x^k; r_k) \leqslant F(x; r_k) \quad \forall x \in G_N^\delta \cap G_L. \tag{6.103}$$

При $x = \bar{x}$ из (6.103) и (6.102) следует существование числа $r(\delta)$ и точки $x^k \in G_N^\delta \cap G_L$ таких, что для $r_k > r(\delta)$ справедливо неравенство

$$F(x^k; r_k) \leqslant F(x; r_k) \quad \forall x \in G_L.$$

Теорема 6.18 позволяет дать оценку, которая может служить обоснованием для выбора критерия окончания счета в алгоритме штрафных функций.

Теорема 6.19. *Пусть выполнены условия теоремы* 6.18 *и точка* z^k *выбрана согласно* (6.97)–(6.99). *Тогда для любого* $\epsilon > 0$ *найдется число*

$r_0(\epsilon)$ *такое, что для всех* k, *при которых* $r_k > r_0(\epsilon)$, *выполняется соотношение*

$$0 \leqslant f(z^k) - f(x^k) \leqslant \epsilon.$$

Доказательство. Так как функция f: $\mathbf{R}^n \to \mathbf{R}$ выпукла, то из (6.97) следует

$$f(z^k) \leqslant \alpha_k f(\bar{x}) + (1 - \alpha_k) f(x^k)$$

и, с учетом (6.100),

$$0 \leqslant f(z^k) - f(x^k) \leqslant (f(\bar{x}) - f(x^k))\alpha_k. \tag{6.104}$$

Для $r_k > r(\delta)$ по теореме 6.15 имеем $x \in G_N^\delta \cap G_L$, т.е. $g_i(x^k) \leqslant \delta \quad \forall i \in I$. Полагая $\gamma = \min\limits_{i \in I_k} \{-g_i(\bar{x})\}$, получим для α_k оценку

$$\alpha_k < \delta/(\gamma + \delta) \tag{6.105}$$

для всех k, при которых $r_k > r(\delta)$.

Так как множество $G_N^\delta \cap G_L$ компактно, а функция f непрерывна, то найдется число K такое, что

$$|f(\bar{x}) - f(x)| \leqslant K \qquad \forall x \in G_N^\delta \cap G_L. \tag{6.106}$$

Таким образом, из (6.104), (6.105), (6.106) при $\epsilon = K\delta/(\gamma + \delta)$ следует что $0 \leqslant f(z^k) - f(x^k) \leqslant \epsilon$ для всех k, при которых $r_k > r(\delta)$.

Полагая $\delta = \epsilon\gamma/(K - \epsilon)$ и $r_0(\epsilon) = r(\epsilon\gamma/(K - \epsilon))$, получаем утверждение теоремы.

Замечание. В случае $S(x) = \max\limits_{i \in I} \{0, g_i(x)\}$ можно выбрать $r_0(\epsilon) = K^2/\epsilon\gamma$. Так как $S(x) = \delta \quad \forall x \in \hat{G}_N^\delta \cap G_L$, то $S(x^2) = \delta$, и поэтому

$$r(\delta) = \frac{|f(\bar{x}) - f(x^1)|}{S(x^2)} \leqslant \frac{K}{\delta} = \frac{K(K - \epsilon)}{\epsilon\gamma} < \frac{K^2}{\epsilon\gamma} = r_0(\epsilon).$$

6.6.4. Метод барьеров. В методе барьеров целевые функции F_k вспомогательных задач H_k выбираются таким образом, чтобы решения H_k всегда лежали во внутренности множества G (см. рис. 6.19). Метод барьеров называют также *методом внутреннего штрафа* в противоположность рассмотренному в пп. 6.6.2 и 6.6.3 *методу внешнего штрафа.*

Для вспомогательных задач H_k предположим, что

$$F_k(x) = F(x; q_k) = f(x) + q_k B(x), \tag{6.107}$$

функция B: $Q \to \mathbf{R}$, $Q \subseteq \mathbf{R}^n$, непрерывна на множестве *"точек Слейтера"*

$$G^< = \{x \in Q: \quad g_i(x) < 0, \quad i \in I\}, \tag{6.108}$$

$\lim\limits_{k \to \infty} B(x^k) = \infty$ для любой последовательности $\{x^k\} \subset G^<$ такой, что

$$x^k \to y \in G \setminus G^<, \qquad \|y\| < \infty. \tag{6.109}$$

От последовательности $\{q_k\}$ потребуем, чтобы

$$q_k > q_{k+1} > 0 \quad \text{для} \quad k = 1, 2, \ldots, \tag{6.110}$$

$$\lim_{k \to \infty} q_k = 0. \tag{6.111}$$

Утверждение о существовании решений x^k вспомогательной задачи

$$H_k\colon \min\{f(x)+q_k B(x)\colon\ x\in G^{<}\cap G_L\} \tag{6.112}$$

содержит следующая теорема.

Теорема 6.20. *Пусть множество $G_N\cap G_L$ компактно и $G^{<}\cap G_L\neq\phi$. Тогда для задач H_k существуют решения x^k, т.е.*

$$F(x^k;\ q_k)=\min_{x\in G^{<}\cap G_L} F(x;\ q_k). \tag{6.113}$$

Доказательство. Пусть $\hat{x}\in G^{<}\cap G_L$ – произвольный фиксированный вектор. Определим

$$M_k=\{x\in\mathbf{R}^n\colon\ x\in G^{<}\cap G_L,\quad F(x;q_k)\leqslant F(\bar{x};\ q_k)\}. \tag{6.114}$$

Для доказательства теоремы достаточно показать, что множество M_k компактно.

Так как $M_k\subseteq G_N\cap G_L$, то M_k ограничено. Поэтому остается удостовериться в замкнутости M_k. Пусть последовательность $\{x^l\}\subset M_k$ сходится к y. Тогда $F(x^l;\ q_k)\leqslant F(\bar{x};\ q_k)$ и

$$F(y;q_k)=\lim_{l\to\infty} F(x^l;\ q_k)\leqslant F(\bar{x};\ q_k). \tag{6.115}$$

Поскольку множество $G_N\cap G_L$ компактно, то $y\in G_N\cap G_L$. Если допустить, что $y\in G_N\setminus G^{<}$, то вследствие (6.109) получим $\lim\limits_{l\to\infty} B(x^l)=\infty$, а это противоречит (6.115). Отсюда следует $y\in G^{<}\cap G_L$ и, в силу (6.115), $y\in M_k$. Таким образом,

$$\inf_{x\in G^{<}\cap G_L} F(x;\ q_k)=\inf_{x\in M_k} F(x;\ q_k)=F(x^k;\ q_k).$$

Алгоритм метода барьеров соответствует по своей структуре алгоритму метода штрафов.

В последующих двух теоремах обсуждается сходимость метода барьеров.

Теорема 6.21. *Пусть $\{x^k\}$ – последовательность, порождаемая методом барьеров. Тогда*

$$f(x^k)\geqslant f(x^{k+1})\geqslant f(x^0),\ k=1,\ 2,\ldots \tag{6.116}$$

Если дополнительно к свойствам (6.108) *и* (6.109) *выполняется условие*

$$B(x)\geqslant 0\quad \forall x\in G^{<}, \tag{6.117}$$

то

$$F(x^k;\ q_k)\geqslant F(x^{k+1};\ q_{k+1}),\quad k=1,\ 2,\ldots, \tag{6.118}$$

$$F(x^k;\ q_k)\geqslant f(x^k)\geqslant f(x^0). \tag{6.119}$$

Доказательство этой теоремы аналогично доказательству теоремы 6.15.

Чтобы обеспечить сходимость метода барьеров, введем дополнительное условие регулярности

$$\inf_{x\in G^{<}\cap G_L} f(x)=\min_{x\in G_N\cap G_L} f(x). \tag{6.120}$$

Это условие выполняется не всегда.

Пример 6.10. Для задачи (рис. 6.22)

$$\min\{-x_1 : g(x_1) \leqslant 0, \quad x_1 \in \mathbf{R}\},$$

$$g(x_1) \triangleq \begin{cases} x_1^2 - 1 & \text{при } x_1 < 1, \\ 0 & \text{при } 1 \leqslant x_1 \leqslant 2, \\ 2x_1 - 4 & \text{при } x_1 > 2 \end{cases}$$

имеем $G^< = (-1, 1)$, $G_N = [-1, 2]$,

$$\inf_{x_1 \in G^<} f(x_1) = -1, \quad \min_{x_1 \in G_N} f(x_1) \triangleq -2.$$

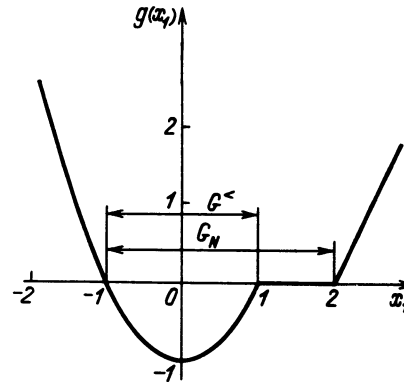


Рис. 6.22.

Для выполнения условия регулярности (6.120) достаточно, например, чтобы (см. [19])

$$\overline{G^< \cap G_L} = G_N \cap G_L. \qquad (6.121)$$

Теорема 6.22. *Пусть последовательность* $\{x^k\}$ *порождена методом барьеров согласно* (6.113). *Предположим, что выполнены условия теоремы* 6.20 *и условие регулярности* (6.120). *Тогда любая предельная точка последовательности* $\{x^k\}$ *является решением задачи* P.

Доказательство. Пусть $\{x^{k_j}\}$ — сходящаяся к точке y подпоследовательность из $\{x^k\}$. Так как множество $G_N \cap G_L$ компактно, то $y \in G_N \cap G_L$. Согласно (6.113)

$$F(x^{k_j}; q_{k_j}) \leqslant F(x; q_{k_j}) \quad \forall x \in G^< \cap G_L. \qquad (6.122)$$

Поскольку функция B ограничена на множестве $G^< \cap G_L$, то

$$\lim_{k_j \to \infty} F(x; q_{k_j}) = f(x)$$

и, с учетом (6.122),

$$\overline{\lim_{k_j \to \infty}} F(x^{k_j}; q_{k_j}) \leqslant f(x) \quad \forall x \in G^< \cap G_L. \qquad (6.123)$$

Выделим два случая.

Случай 1: $y \in G_N \setminus G^<$. Тогда в силу (6.109) получаем $B(x^{k_j}) \to \infty$ при $k_j \to \infty$ и, таким образом, $q_{k_j} B(x^{k_j}) > 0$ для достаточно больших k_j, т.е. $\underline{\lim}_{k_j \to \infty} q_{k_j} B(x^{k_j}) \geqslant 0$.

Случай 2: $y \in G^<$. Из непрерывности функции B на множестве G и соотношения $q_{k_j} \to 0$ следует

$$\lim_{k_j \to \infty} q_{k_j} B(x^{k_j}) = 0.$$

Итак, в обоих случаях $\underline{\lim}_{k_j \to \infty} q_{k_j} B(x^{k_j}) \geqslant 0$, а тем самым и

$$\underline{\lim}_{k_j \to \infty} F(x^{k_j}; q_{k_j}) \geqslant \underline{\lim}_{k_j \to \infty} f(x^{k_j}) = f(y).$$

Отсюда и из (6.123) вытекает

$$f(y) \leqslant f(x) \quad \forall x \in G^< \cap G_L,$$

и, в силу (6.120),

$$f(y) \leqslant f(x) \quad \forall x \in G_N \cap G_L.$$

Поэтому последовательность $\{f(x^{k_j})\}$ сходится, монотонно убывая к $f(x^0)$. Для соответствующей последовательности $\{F(x^{k_j}; q_{k_j})\}$ это выполняется лишь в случае $B(x) \geqslant 0 \quad \forall x \in G^<$, как было показано в теореме 6.21.

Если множество G_N задано в виде

$$G_N = \{x \in \mathbf{R}^n: \ g_i(x) \leqslant 0, \quad i \in I\},$$

то для функции B можно привести следующие реализации:

$$1) \ B(x) = \frac{1}{\max\limits_{i \in I} g_i(x)}; \tag{6.124}$$

$$2) \ B(x) = \sum_{i=1}^{m} b(-g_i(x)), \tag{6.125}$$

где

функция b: $\mathbf{R} \to \mathbf{R}$ непрерывна при $t > 0$, (6.126)

$$\lim_{t \to +0} b(t) = +\infty. \tag{6.127}$$

Примеры функции b:

$$b = b_1(t) = t^{-p} \quad (p > 0), \quad b = b_2(t) = -\ln t.$$

6.6.5. Метод барьеров для задач выпуклого программирования. Для получения оценок погрешности в задачах выпуклого программирования при $G_L = \mathbf{R}^n$ в этом пункте привлекаются теоремы двойственности. Согласно (5.29) задача, двойственная к задаче

$$P: \min\{f(x): \ x \in G_N\}, \qquad G_N = \{x \in \mathbf{R}^n: \ g_i(x) \leqslant 0, \quad i \in I\},$$

имеет вид

$$\max\{f(x) + \sum_{i \in I} y_i g_i(x): \ \nabla f(x) + \sum_{i \in I} y_i \nabla g_i(x) = 0, \quad y \geqslant 0\}. \tag{6.128}$$

Установим связь между методом барьеров и решениями двойственной задачи (6.128).

Теорема 6.23. *Пусть функции f, g_i, $i \in I$, выпуклы и непрерывно дифференцируемы, функция $F(x; q_k)$ образована в соответствии с* (6.125)–(6.127), *при этом функция b непрерывно дифференцируема и*

$$-\infty < db/dt < 0 \quad (t > 0). \tag{6.129}$$

Предположим, далее, что последовательность $\{x^k\}$ порождена методом барьеров, множество G_N компактно и $G^< \neq \phi$. Справедливы утверждения.

1) *Последовательность $\{(x^k, y^k)'\}$, где*

$$y_i^k = -q_k \left.\frac{db}{dt}\right|_{t = -g_i(x^k)}, \quad i \in I, \tag{6.130}$$

является последовательностью допустимых в двойственной задаче (6.128) *точек; при этом*

$$f(x^k) + \sum_{i=1}^{m} y_i^k g_i(x^k) \leqslant f(x^0) \leqslant f(x^k). \tag{6.131}$$

2) *Любая предельная точка последовательности* $\{(x^k, y^k)'\}$ *будет решением задачи* (6.128).

Доказательство 1. Для того чтобы точка x^k была решением задачи (6.112) при $G_L = \mathbf{R}^n$, необходимо выполнение условия

$$\nabla_x F(x^k; q_k) = \nabla f(x^k) - q_k \sum_{i=1}^{m} \frac{db}{dt}\bigg|_{t=-g_i(x^k)} \nabla g_i(x^k) = 0.$$

Если мы выберем y_i^k согласно (6.130), то будем иметь

$$\nabla f(x^k) + \sum_{i \in I} y_i^k \nabla g_i(x^k) = 0, \tag{6.132}$$

$$y_i^k \geqslant 0, \quad i \in I. \tag{6.133}$$

Поэтому точка $(x^k, y^k)'$ допустима в задаче (6.128). Отсюда с учетом слабой теоремы двойственности сразу же следует соотношение (6.131).

2. Пусть x^* – предельная точка последовательности $\{x^k\}$ и $\lim\limits_{k_j \to \infty} x^{k_j} = x^*$. Для $i \notin I_0^* = \{i \in I: g_i(x^*) = 0\}$ из непрерывности функций g_i, $i \in I$, и db/dt следует

$$-\infty < \lim_{k_j \to \infty} \frac{db}{dt}\bigg|_{t=-g_i(x^{k_j})} = \frac{db}{dt}\bigg|_{t=-g_i(x^*)} \leqslant 0.$$

Учитывая, что $\lim\limits_{k_j \to \infty} q_{k_j} = 0$, получим

$$\lim_{k_j \to \infty} \left(-q_{k_j} \frac{db}{dt}\bigg|_{t=-g_i(x^{k_j})} \right) = \lim_{k_j \to \infty} y_i^{k_j} = 0, \quad i \notin I_0^*, \tag{6.134}$$

и, таким образом, в силу ограниченности $g_i(x^{k_j})$ ($x^{k_j} \in G_N$, а G_N компактно),

$$\lim_{k_j \to \infty} y_i^{k_j} g_i(x^{k_j}) = 0, \quad i \notin I_0^*. \tag{6.135}$$

Так как функции $f, g_i, i \in I$, выпуклы и непрерывно дифференцируемы, то условие (6.132) необходимо и достаточно для того, чтобы при $k = k_j$ выполнялось

$$f(x) + \sum_{i \in I} y_i^{k_j} g_i(x) \geqslant f(x^{k_j}) + \sum_{i \in I} y_i^{k_j} g_i(x^{k_j}) \quad \forall x \in \mathbf{R}^n. \tag{6.136}$$

В силу соотношений (6.129), (6.130) и $q_{k_j} > 0$ имеем $y_i^{k_j} > 0$ для $i \notin I_0^*$ и всех k_j.

Пусть $\mu_i^{k_j} = y_i^{k_j} \left(\sum_{i \in I} y_i^{k_j} \right)^{-1}$. Тогда после деления обеих частей неравенства (6.136) на $\sum_{i \in I} y_i^{k_j}$ для $x = \bar{x} \in G^<$ получим

$$\sum_{i \in I} \mu_i^{k_j} (g_i(\bar{x}) - g_i(x^{k_j})) = \frac{f(x^{k_j}) - f(\bar{x})}{\sum_{i \in I} y_i^{k_j}} . \tag{6.137}$$

Если допустить, что $\lim_{k_j \to \infty} \sum_{i \in I} y_i^{k_j} = \infty$, то из соотношений (6.135) и

$$\lim_{k_j \to \infty} g_i(x^{k_j}) = 0, \quad i \in I_0^*, \tag{6.138}$$

будет вытекать неравенство $\lim_{k_j \to \infty} \sum_{i \in I} \mu_i^{k_j} g_i(\bar{x}) \geqslant 0$, а это противоречит соотношениям $g_i(\bar{x}) < 0$, $\mu_i^{k_j} \geqslant 0 \ \ \forall i \in I$ и $\sum_{i \in I} \mu_i^{k_j} = 1$.

Следовательно, $\lim_{k_j \to \infty} \sum_{i \in I} y_i^{k_j} < \infty$, и в силу (6.133) и (6.134) имеем

$$\lim_{k_j \to \infty} y_i^{k_j} < \infty \quad \forall i \in I_0^* . \tag{6.139}$$

Из соотношений (6.138) и (6.139) получаем

$$\lim_{k_j \to \infty} y_i^{k_j} g_i(x^{k_j}) = 0 \quad \forall i \in I_0^* .$$

Отсюда и из (6.135) следует, что

$$\lim_{k_j \to \infty} \left\{ f(x^{k_j}) + \sum_{i \in I} y_i^{k_j} g_i(x^{k_j}) \right\} = f(x^*) . \tag{6.140}$$

По теореме 6.22 $f(x^*) = f(x^0)$ и в силу (6.131) из (6.140) вытекает, что любая предельная точка последовательности $\{(x^k, y^k)'\}$ с y^k из (6.130) является решением двойственной задачи (6.128).

Замечание. Соотношение (6.131) может служить основанием для критерия окончания счета, так как

$$\sum_{i \in I} y_i^k g_i(x^k) \leqslant f(x^0) - f(x^k) \leqslant 0.$$

В случае $b(t) = -\ln t$ имеем, в частности,

$$\sum_{i \in I} y_i^k g_i(x^k) = \sum_{i \in I} - q_k \frac{g_i(x^k)}{g_i(x^k)} = -mq_k .$$

6.6.6. Смешанный метод штрафов и барьеров. Методы штрафов и барьеров могут применяться также сообща в форме "смешанного" метода. Мы ограничимся кратким описанием данного подхода (см. общее изложение в работе [27]).

Пусть допустимое множество задачи P представимо в виде

$$G = G_B \cap G_S \cap G_L,$$

где

$$G_B = \{ x \in \mathbf{R}^n : g_i(x) \leqslant 0, \ i = 1, \ldots, m_1 \},$$

$$G_S = \{ x \in \mathbf{R}^n : g_i(x) \leqslant 0, \ i = m_1 + 1, \ldots, m \}.$$

Вспомогательные функции F строятся таким образом, чтобы они были барьерными относительно множества G_B и штрафными относительно G_S:

$$F(x; r_k) = f(x) + r_k S(x) + \frac{1}{r_k} B(x).$$

Пусть множество G компактно, G_B и G_S замкнуты, $G_B^< \cap G_S \cap G_L \neq \phi$, выполнено условие регулярности

$$\inf_{x \in G_B^< \cap G_S \cap G_L} f(x) = \min_{x \in G_B \cap G_S \cap G_L} f(x),$$

функция $F(x; r_k)$ непрерывна на $G_B^< \cap G_L$, функция S обладает свойствами (6.81) и (6.82) (с заменой G_N на G_S), функция B удовлетворяет условиям (6.108) и (6.109) (с заменой G_N и $G_N^<$ соответственно на G_B и $G_B^<$) и $\{r_k\}$ – строго возрастающая последовательность положительных чисел. При этих предположениях может быть доказано утверждение, аналогичное теоремам 6.16 и 6.22.

В заключение приведем следующий пример.

Таблица 6.5

	q_k	r_k	x_1^k	x_2^k
Случай 1	0,1	10	0,685378	1,266645
	0,01	100	0,704638	1,399240
	0,001	1000	0,706857	1,412714
	0,0001	10000	0,707082	1,414068
	0,00001	100000	0,707104	1,414199
Случай-2	0,1	10	0,713012	1,451244
	0,01	100	0,707727	1,417950
	0,001	1000	0,707169	1,414588
	0,0001	10000	0,707113	1,414251
	0,00001	100000	0,707107	1,414217
Случай 3	0,1	–	0,796035	1,266349
	0,01	–	0,715948	1,401680
	0,001	–	0,707982	1,412963
	0,0001	–	0,707194	1,414088
	0,00001	–	0,707116	1,414201
Случай 4	–	10	0,652590	1,330421
	–	100	0,700882	1,406738
	–	1000	0,706544	1,413339
	–	10000	0,707051	1,414126
	–	100000	0,707101	1,414205

Пример 6.11 [36]. В задаче

$$\max\{x_2:\ x=(x_1,x_2)'\in\mathbf{R}^2;\ 2x_1-x_2\geqslant 0,\ 1-x_1x_2\geqslant 0\},\qquad (6.141)$$

как легко видеть, максимум достигается в точке $x^0=(\sqrt{2}/2,\ \sqrt{2})'$. В x^0 оба ограничения активны. Последующие четыре случая выбора вспомогательных задач показывают, каким образом x^k приближаются к решению x^0, если ограничения поглощаются штрафными или барьерными функциями. Обозначим зависящие от x, r_k и q_k вспомогательные функции через $F_k(x)$. Итерационные точки x^k находятся как решения системы

$$\nabla F_k(x)=0,\quad k=1,2,\ldots \qquad (6.142)$$

Случай 1: $F_k(x)=x_2+q_k\ln(2x_1-x_2)+q_k\ln(1-x_1x_2)$. Из (6.142) следует

$$\frac{2q_k}{2x_1^k-x_2^k}-\frac{q_kx_2^k}{1-x_1^kx_2^k}=0,\quad 1-\frac{q_k}{2x_1^k-x_2^k}-\frac{q_kx_1^k}{1-x_1^kx_2^k}=0.$$

Преобразовывая, получим

$$x_1^k=2q_kx_2^k/(2(q_k+x_2^k)),$$

$$2(x_2^k)^3+5q_k(x_2^k)^2-4x_2^k+2q_k=0,\quad k=1,2,\ldots$$

Случай 2: $F_k(x)=x_2-r_k(\min\{0,2x_1-x_2\})^2-r_k(\min\{0,1-x_1x_2\})^2$. Из (6.142) после преобразований получим

$$x_1^k=\frac{3x_2^k}{4+(x_2^k)^2},$$

$$-2r_k(x_2^k)^5+(x_2^k)^4-16r_k(x_2^k)^3+8(x_2^k)^2+40r_kx_2^k+16=0,\quad k=1,2,\ldots$$

Случай 3: $F_k(x)=x_2+q_k\ln(2x_1-x_2)-r_k(\min\{0,1-x_1x_2\})^2$. Полагая для упрощения $q_k=1/r_k$, из (6.142) получим следующее:

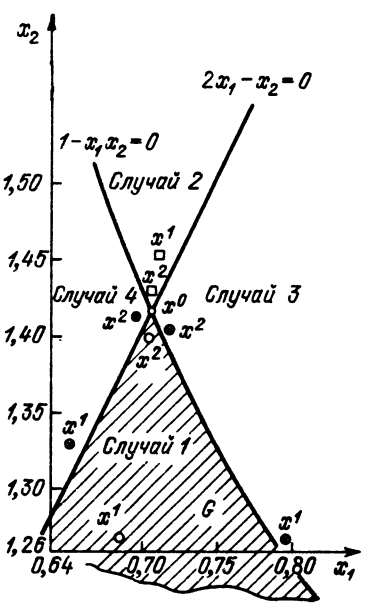


Рис. 6.23.

$$x_1^k=\frac{x_2^k(x_2^k+2q_k)}{x_2^k-q_k},$$

$$-r_k(x_2^k)^5-5(x_2^k)^4+(r_k-2q_k)(x_2^k)^3-$$

$$-8q_kx_2^k+2q_k+q_k^2=0,\quad k=1,2,\ldots$$

Случай 4: $F_k(x)=x_2-r_k(\min\{0,2x_1-x_2\})^2+q_k\ln(1-x_1x_2)$. Полагая снова

$$q_k=\frac{1}{r_k},$$

из (6.142) получим

$$x_1^k=(2-q_kx_2^k)/(2(x_2^k+q_k)),$$

$$-r_k(x_2^k)^4-3(x_2^k)^3+2q_k(x_2^k)^2+$$

$$+(q_k^2-8)x_2^k+8r_k=0,\quad k=1,2,\ldots$$

Соответствующие результаты счета для рассматриваемых четырех случаев содержатся в табл. 6.5.

Рис. 6.23 иллюстрирует характер приближения членов последовательности $\{x^k\}$ к решению x^0 исходной задачи.

6.7. Методы секущих плоскостей

Рассмотрим задачу

$$\min\{c'x:\ x \in G\}, \tag{6.143}$$

где $G = \{x \in \mathbf{R}^n:\ g_i(x) \leqslant 0,\ i \in I\}$, $I = \{1, \ldots, m\}$, $g_i:\ \mathbf{R}^n \to \mathbf{R}$ – выпуклые функции.

Применение линейной функции в задаче (6.143), как уже отмечалось в § 6.4, не представляет ограничения общности по сравнению с постановкой P_1.

Положим

$$g(x) = \max_{i \in I} g_i(x), \tag{6.144}$$

$$D_\tau(x^k, y^k) = \{z \in \mathbf{R}^n:\ (x^k - y^k)'z \geqslant \tau \| x^k - y^k \| \, \| z \|\}. \tag{6.145}$$

При выполнении условий

$V1$: множество G ограничено и содержится в многограннике M_1,

$V2$: $G^< \neq \phi$, где $G^< = \{x \in \mathbf{R}^n:\ g_i(x) < 0 \quad \forall i \in I\}$,

мы покажем сходимость следующего алгоритма:

Шаг 0. Заданы $M_1 \supseteq G$ и $\tau \in (0, 1]$. Положить $k = 1$.

Шаг 1. Определить решение x^k задачи

$$P_L^k:\ \min\{c'x:\ x \in M_k\}.$$

Шаг 2. Если $x^k \in G$, то остановиться. В противном случае выбрать векторы $y^k \in M_k \backslash G^<$ и $z^k \in \partial g(y^k)$ такие, чтобы выполнялось

$$\partial g(y^k) \cap D_\tau(x^k, y^k) \neq \phi.$$

Шаг 3. Образовать множества

$$E^{(k)} = \{x \in \mathbf{R}^n:\ g(x^k) + (x - y^k)'z^k \leqslant 0\},$$

$$M_{k+1} = M_k \cap E^{(k)}.$$

Шаг 4. Положить $k := k + 1$ и перейти к шагу 1.

Вырабатываемые данным методом точки x^k и соответствующие многогранники M_k имеют следующие свойства.

Теорема 6.24. *Пусть x^0 – решение задачи* (6.143). *Справедливы утверждения*

1) $G \subseteq M_k \quad \forall k \in \mathbf{N}$;

2) $c'x^k \leqslant c'x^0 \quad \forall k \in \mathbf{N}$.

Доказательство. 1. Пусть $x \in G$. Согласно шагу 2 $z^k \in \partial g(y^k)$, т.е. $g(y^k) + (x - y^k)'z^k \leqslant g(x)$. С учетом неравенства $g(x) \leqslant 0$ имеем

$$g(y^k) + (x - y^k)'z^k \leqslant 0 \quad \forall k \in \mathbf{N}.$$

Итак, $x \in E^{(k)} \ \forall k \in \mathbf{N}$, и, следовательно, $G \subseteq E^{(k)} \quad \forall k \in \mathbf{N}$.

Из определения шага 3 и условия $V1$ следует $G \subseteq M_k \quad \forall k \in \mathbf{N}$.

2. Второе утверждение непосредственно вытекает из первого.

При доказательстве сходимости метода нам понадобится следующая теорема.

Теорема 6.25. *Если* $y \in M_1 \setminus G^<$, *то* $0 \notin \partial g(y)$.

Доказательство. Для любого $z \in \partial g(y)$ имеем

$$g(x) \geqslant g(y) + (x - y)'z \qquad \forall x \in \mathbf{R}^n.$$

Так как $y \in M_1 \setminus G^<$, то $g(y) \geqslant 0$, и поэтому, с одной стороны,

$$g(x) \geqslant (x - y)'z \qquad \forall x \in \mathbf{R}^n,$$

а с другой, в силу условия $V2$, существует вектор $\bar{x} \in G^<$ такой, что

$$0 > g(\bar{x}) \geqslant (\bar{x} - y)'z.$$

Отсюда следует $z \neq 0$.

При исследовании свойств сходимости метода будем различать два случая.

Случай 1: после конечного числа шагов алгоритм останавливается на шаге 2 при $k = k_0$, так что $x^{k_0} \in G$. Тогда по теореме 6.24 имеем $c'x^{k_0} \leqslant c'x^0$, $x^{k_0} \in G$. Тем самым x^{k_0} является решением задачи (6.143).

Случай 2: алгоритм вырабатывает бесконечную последовательность $\{x^k\}$ и $x^k \notin G \qquad \forall k \in \mathbf{N}$.

Этот случай мы подробнее исследуем ниже.

Теорема 6.26. *Пусть выполнено условие* $V1$ *и* $x^k \notin G \ \forall k \in \mathbf{N}$. *Тогда любая предельная точка* y^* *последовательности* $\{y^k\}$ *принадлежит множеству* G. *Если* $y^k = x^k \ \forall k \in \mathbf{N}$, *то* y^* *– решение задачи* (6.143).

Доказательство. Из условия $V1$ следует существование предельной точки y^* последовательности $\{y^k\}$. Пусть $\{y^{k_l}\}$ – сходящаяся подпоследовательность из $\{y^k\}$ и $\lim\limits_{k_l \to \infty} y^{k_l} = y^*$.

Допустим, что $y^* \notin G$. Тогда $g(y^*) > 0$, и вследствие непрерывности функции g: $\mathbf{R}^n \to \mathbf{R}$ и сходимости последовательности $\{y^{k_l}\}$ существует номер l_0 такой, что

$$g(y^{k_l}) \geqslant \frac{1}{2} g(y^*) \qquad \forall l \geqslant l_0. \tag{6.146}$$

Кроме того, из описания шага 3 следует, что

$$y^{k_{l+1}} \in M_{k_{l+1}} \setminus G^< \subset M_{k_{l+1}} \subset M_{k_l} \cap E^{(k_l)},$$

а поэтому

$$0 \geqslant g(y^{k_l}) + (y^{k_{l+1}} - y^{k_l})'z^{k_l}.$$

Согласно теореме 2.80 субдифференциал функции g является компактным множеством. Вследствие этого существует число $K > 0$ такое, что $\|z^{k_l}\| \leqslant K \qquad \forall z^{k_l} \in \partial g(y^{k_l})$. Поэтому

$$0 \geqslant g(y^{k_l}) - \|y^{k_{l+1}} - y^{k_l}\| \cdot K.$$

В силу (6.146) имеем

$$0 \geqslant \frac{1}{2} g(y^*) - \|y^{k_{l+1}} - y^{k_l}\| \cdot K \qquad \forall l \geqslant l_0,$$

и, следовательно,

$$\| y^{k_{l+1}} - y^{k_l} \| \geqslant g(y^*)/(2K) > 0 \quad \forall l \geqslant l_0.$$

Последнее противоречит сходимости последовательности $\{y^{k_l}\}$. Таким образом, $y^* \in G$.

Если $y^k = x^k \ \ \forall k \in \mathbf{N}$, то из второго утверждения теоремы 6.24 сразу же следует, что y^* – решение задачи (6.143).

В общем случае $y^k \neq x^k$ справедлива следующая теорема.

Теорема 6.27. *Пусть выполнены условия* $V1$ *и* $V2$, $x^k \notin G \ \ \forall k \in \mathbf{N}$ *и* $y^k \neq x^k \ \ \forall k \in \mathbf{N}$. *Тогда любая предельная точка* x^* *последовательности* $\{x^k\}$ *будет решением задачи* (6.143).

Доказательство. Согласно условию $V1$ последовательности $\{x^k\}$ и $\{y^k\}$ имеют сходящиеся подпоследовательности $\{x^{k_l}\}$ и $\{y^{k_l}\}$ соответственно. Пусть $\lim\limits_{k_l \to \infty} x^{k_l} = x^*$ и $\lim\limits_{k_l \to \infty} y^{k_l} = y^*$. Так как $x^{k_{l+1}} \in E^{(k_l)}$, то

$$0 \geqslant g(y^{k_l}) + (x^{k_{l+1}} - y^{k_l})' z^{k_l} \geqslant (x^{k_{l+1}} - y^{k_l})' z^{k_l}.$$

Поэтому из разложения

$$(x^{k_l} - x^{k_{l+1}})' z^{k_l} = (x^{k_l} - y^{k_l})' z^{k_l} - (x^{k_{l+1}} - y^{k_l})' z^{k_l}$$

следует неравенство

$$(x^{k_l} - x^{k_{l+1}})' z^{k_l} \geqslant (x^{k_l} - y^{k_l})' z^{k_l}. \tag{6.147}$$

Отсюда получаем

$$\| x^{k_l} - x^{k_{l+1}} \| \, \| z^{k_l} \| \geqslant (x^{k_l} - x^{k_{l+1}})' z^{k_l} \geqslant (x^{k_l} - y^{k_l})' z^{k_l}.$$

В силу $z^{k_l} \in D_\tau(x^{k_l}, y^{k_l})$ имеем

$$(x^{k_l} - y^{k_l})' z^{k_l} \geqslant \tau \| x^{k_l} - y^{k_l} \| \, \| z^{k_l} \| \geqslant 0,$$

следовательно,

$$\| x^{k_l} - x^{k_{l+1}} \| \, \| z^{k_l} \| \geqslant \tau \| x^{k_l} - y^{k_l} \| \, \| z^{k_l} \|$$

и, с учетом теоремы 6.25,

$$\| x^{k_l} - x^{k_{l+1}} \| \geqslant \tau \| x^{k_l} - y^{k_l} \| \geqslant 0.$$

Поскольку $\{x^{k_l}\}$ сходится и τ положительно, имеем $\| x^{k_l} - y^{k_{l+1}} \| \to 0$ при $k_l \to \infty$. Таким образом, $x^* = y^*$. В силу теорем 6.26 и 6.24 вектор $x^* \in G$ является решением задачи (6.143).

Предыдущие рассуждения показывают, что при выборе $y^k = x^k$ свойство $\partial g(y^k) \cap D_\tau(x^k, y^k) \neq \phi$ выполняется всегда, потому что

$$D_\tau(x^k, x^k) = \mathbf{R}^n.$$

В этом частном случае сходимость получается из теоремы 6.26. Поэтому здесь выполнения условия $V2$ не требуется.

Напротив, в общем случае $y^k \in M_k \backslash G^<$, $y^k \neq x^k$, условие $\partial g(y^k) \cap D_\tau(x^k, y^k) \neq \phi$ при $\tau > 0$ необходимо на шаге 2, как это показывает доказательство теоремы 6.27.

Теперь возникает вопрос, всегда ли это условие выполнимо на шаге 2. Покажем, что при подходящем выборе $y^k \in M_k \backslash G^<$ всегда существует число $\tau > 0$ такое, что $\partial g(y^k) \subset D_\tau(x^k, y^k)$.

Для этого положим

$$G_\gamma^< = \{ x \in \mathbf{R}^n : g(x) \leqslant -\gamma < 0; \ \gamma \text{ фиксировано} \}$$

и обозначим через $C(G_\gamma^< \cup \{x^k\})$ выпуклую оболочку множеств $G_\gamma^<$ и $\{x^k\}$.

Теорема 6.28. *Пусть* $x^k \notin G$, $G_\gamma^< \neq \phi$ *и* $y^k \in C(G_\gamma^< \cup \{x^k\}) \backslash G^<$. *Тогда существует число* $\tau > 0$, *для которого*

$$\partial g(y^k) \subseteq D_\tau(x^k, y^k) \qquad \forall k \geqslant 1.$$

Доказательство. Пусть $k \geqslant 1$ – произвольное натуральное число. В силу $x^k \notin G$ и $G_\gamma^< \neq \phi$ для точки $a^k \in G_\gamma^<$, удовлетворяющей неравенству $g(x) \leqslant -\gamma < 0$, всегда можно выбрать y^k так, чтобы выполнялось

$$a^k - y^k = \frac{\mu_k}{1 - \mu_k}(y^k - x^k), \qquad 1 > \mu_k > 0. \tag{6.148}$$

Согласно условию $V1$ найдется число $K_1 > 0$ такое, что

$$\| y^k - a^k \| \leqslant K_1, \tag{6.149}$$

а по теореме 2.80 для функции g существует число $K_2 > 0$, при котором

$$\| z^k \| \leqslant K_2 \qquad \forall z^k \in \partial g(y^k). \tag{6.150}$$

Тогда

$$0 < \gamma < -g(a^k) \leqslant -g(y^k) - (a^k - y^k)' z^k \leqslant (y^k - a^k)' z^k.$$

Отсюда следует, что

$$\gamma \frac{\| y^k - a^k \| \| z^k \|}{\| y^k - a^k \| \| z^k \|} \leqslant (y^k - a^k)' z^k,$$

и в силу (6.149) и (6.150)

$$\frac{\gamma}{K_1 K_2} \| y^k - a^k \| \| z^k \| \leqslant (y^k - a^k)' z^k.$$

Поэтому, применяя (6.148), получаем для $\tau = \gamma/(K_1 K_2)$ неравенство

$$(x^k - y^k)' z^k \geqslant \tau \| x^k - y^k \| \| z^k \|.$$

Таким образом, если $z^k \in \partial g(y^k)$, то $z^k \in D_\tau(x^k, y^k)$ $\forall k \geqslant 1$. Так как $\partial g(y^k) \neq \phi$, окончательно имеем

$$\phi \neq \partial g(y^k) \subseteq D_\tau(x^k, y^k) \qquad \forall k \geqslant 1.$$

В зависимости от выбора и конструкции векторов y^k и z^k из рассмотренного выше общего метода секущих плоскостей вытекает ряд частных методов.

Если потребовать, чтобы в задаче (6.143) функции $g_i\colon \mathbf{R}^n \to \mathbf{R}$, $i \in I$, были непрерывно дифференцируемы, то z^k можно выбирать в виде

$$z^k = \sum_{i \in I_k} \lambda_i \nabla g_i(y^k), \qquad I_k = \{i \in I\colon g(y^k) = g_i(y^k)\},$$

$$\lambda_i \geqslant 0 \quad \forall i \in I_k; \qquad \sum_{i \in I_k} \lambda_i = 1.$$

При этом получаем $z^k \in \partial g(y^k)$. Действительно, в силу выпуклости функций g_i для $i \in I_k$ выполняется

$$g(x) \geqslant g_i(x) \geqslant g_i(y^k) + (x - y^k)' \nabla g_i(y^k). \tag{6.151}$$

Умножая неравенства (6.151) на $\lambda_i \geqslant 0$ и складывая их, с учетом соотношения $\sum_{i \in I_k} \lambda_i = 1$ имеем

$$g(x) \geqslant g(y^k) + (x - y^k)' \Big[\sum_{i \in I_k} \lambda_i \nabla g_i(y^k)\Big],$$

а потому

$$z^k = \sum_{i \in I_k} \lambda_i \nabla g_i(y^k) \in \partial g(y^k).$$

В методе секущих плоскостей Кэлли [44] шаг 2 модифицируется следующим образом.

Ш а г $2'$. Если $x^k \in G$, то остановиться. В противном случае выбрать $y^k = x^k$ и $z^k = \nabla g_i(y^k)$ для некоторого $i \in I_k$.

Частными примерами метода секущих плоскостей являются методы Клейбома [45] и Топкиса [80].

В методе Клейбома выбирается точка $\bar{x} \in G_\gamma^<$, и роль шага 2 выполняет

Ш а г $2''$. Если $\bar{x} \in G$, то остановиться. В противном случае вычислить $y^k = (1 - \lambda_k)\bar{x} + \lambda_k x^k$, где λ_k – решение задачи

$$\max\{\lambda\colon (1 - \lambda)\bar{x} + \lambda x^k \in G, \quad 0 < \lambda < 1\},$$

и выбрать $z^k \in \partial g(y^k)$.

При таком выборе y^k гиперплоскости $E^{(k)}$ будут опорными к множеству G в граничной точке y^k. Очевидно, что

$$y_k(1 - \lambda_k)\bar{x} + \lambda_k x^k \in C(G_\gamma^< \cup \{x^k\}) \backslash G^<.$$

По теореме 6.28 отсюда следует

$$\partial g(y^k) \cap D_\tau(x^k, y^k) \neq \phi.$$

Так как $y^k \in G$, то в методе Клейбома для $x^0 \in G^0$ имеет место соотношение

$$f(x^k) \leqslant f(x^0) \leqslant f(y^k) \qquad \forall k \in \mathbf{N}.$$

Таким образом, разность $f(y^k) - f(x^k)$ может служить основанием для критерия окончания счета.

К предыдущим рассуждениям добавим несколько замечаний.

З а м е ч а н и я. 1. Требуемое в условии $V1$ многогранное множество M_1 может быть найдено с помощью соответствующих оценок или построено некоторым конечным методом с применением начальной точки, недопустимой в задаче (6.143).

Таблица 6.6.

k	Шаг 1		Шаг 2		
	M_k	Решение P_L^k	$x^k \notin G$	$z^k = \nabla g_i(x^k)$ для некоторого $i \in I_k$	$E^{(k)}$
1	$0 \leqslant x_1 \leqslant 2$ $0 \leqslant x_2 \leqslant 2$	$x^1 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$	да	$\nabla g_2(x^1) = \begin{pmatrix} 4 \\ 4 \end{pmatrix}$	$4x_1 + 4x_2 - 12 \leqslant 0$
2	$M_1 \cap E^{(1)}$	$x^2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$	да	$\nabla g_2(x^2) = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$	$4x_1 + 2x_2 - 9 \leqslant 0$
3	$M_2 \cap E^2$	$x^3 = \begin{pmatrix} 2 \\ \frac{1}{2} \end{pmatrix}$	да		$4x_1 + x_2 - 8\frac{1}{4} \leqslant 0$
4	$M_3 \cap E^{(3)}$	$x^4 = \begin{pmatrix} \frac{15}{8} \\ \frac{4}{3} \end{pmatrix}$	да		
5	– – – –	– – – – – –	– – –	– – – – – – –	– – – – – – –

2. Так как при решении линейных вспомогательных задач P_L^k на каждом итерационном шаге добавляется новое линейное ограничение, то целесообразно вместо P_L^k решать соответствующую двойственную задачу.

3. Существуют методы секущих плоскостей, в которых максимальное число линейных ограничений в задачах P_L^k не превосходит некоторой верхней границы (см., например, работу Топкиса [80]).

4. При использовании методов секущих плоскостей может быть выгодной комбинация различных частных методов. Так, метод Клейбома [45] вначале очень быстро приближается к решению, но потом становится неэффективным из-за определения вспомогательных точек y^k. Поэтому на определенном этапе следует перейти к методу Кэлли.

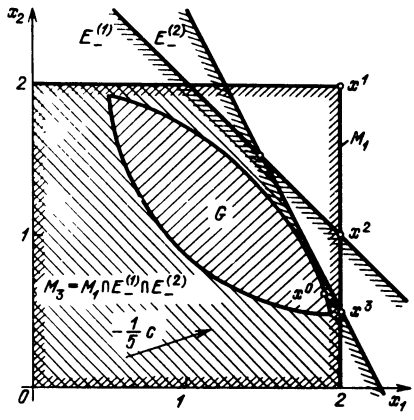


Рис. 6.24.

Для иллюстрации работы метода Кэлли приведем следующий пример.

Пример 6.12. Найти (см. рис. 6.24)

$$\min\{c'x = -3x_1 - x_2 : x \in G \subseteq \mathbf{R}^2\},$$

где

$$G = \{ x \in \mathbf{R}^2 : g_1(x) = (x_1 - 2)^2 + (x_2 - 2)^2 - 9/4 \leqslant 0,$$

$$g_2(x) = x_1^2 + x_2^2 - 4 \leqslant 0\};$$

$$M_1 = \{x \in \mathbf{R}^2 : 0 \leqslant x_1 \leqslant 2,$$

$$0 \leqslant x_2 \leqslant 2\}.$$

Несколько шагов алгоритма представлены в табл. 6.6.

6.8. Методы решения и точечно-множественные отображения

Рассмотренные в § 6.2, 6.3 методы решения основывались на различных идеях. Соответственно исследование сходимости для каждого из обсуждаемых алгоритмов надо было проводить отдельно. Зангвиллом [86] на основе понятия точечно-множественного отображения был предложен общий алгоритм, а также были сформулированы теоремы сходимости для определенных классов алгоритмов одинаковой структуры. Вследствие этого стало возможным унифицировать и частично упростить рассмотрение сходимости методов решения. Кроме того, с помощью этого общего подхода отчетливее проявляется структура конкретных алгоритмов и возникает систематизирующий взгляд на построение методов.

В § 6.1 было введено понятие алгоритма (q, G_1) с помощью точечно-точечного отображения q: $X \to X$. Теперь для решения задачи

$$P: \min \{f(x): \ x \in G\},$$

f: $\mathbf{R}^n \to \mathbf{R}$, $G \subseteq X \subseteq \mathbf{R}^n$, мы определим алгоритм в более общем смысле, применяя точечно-множественное отображение Q: $X \to P(X)$ *), следующим образом:

$$\begin{aligned} &x^{k+1} \in Q(x^k) \quad \forall k \in \mathbf{N}, \\ &x^1 \in G_1, \quad G \subseteq G_1 \subseteq X, \quad x^k \in X \subseteq \mathbf{R}^n \quad \forall k \in \mathbf{N}. \end{aligned} \tag{6.152}$$

При формулировке теорем сходимости для алгоритма (Q, G_1) нам понадобятся определенные свойства непрерывности отображения Q, на которых мы вначале и остановимся. Пусть $X \subseteq \mathbf{R}^n$, $Y \subseteq \mathbf{R}^n$, F: $X \to P(Y)$ — точечно-множественное отображение, $F(X) \neq \phi \ \forall x \in X$.

Отображение F назовем *полунепрерывным сверху* **) (сокращенно: пн.св.) в точке $x^0 \in X$, если справедливо соотношение

$$\left.\begin{aligned} &x^k \to x^0, \quad y^k \to y^0 \\ &y^k \in F(x^k) \end{aligned}\right\} \Rightarrow y^0 \in F(x^0). \tag{6.153}$$

Отображение F называется *полунепрерывным снизу* (сокращенно: пн. сн.) в точке $x^0 \in X$, если имеет место соотношение

$$\left.\begin{aligned} &x^k \to x^0 \\ &y^0 \in F(x^0) \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} &\exists \{y^k\}: \ \lim_{k \to \infty} y^k = y^0, \\ &\exists k_0: \ y^k \in F(x^k) \quad \forall k \geqslant k_0. \end{aligned}\right. \tag{6.154}$$

Если отображение F пн.св. (пн.сн.) в каждой точке множества X, то оно называется пн.св. (пн. сн.) на множестве X. Если отображение F пн. св. и пн.сн., то оно называется непрерывным.

Пример 6.13. На рис. 6.25, *а*, *б* представлены графики двух точечно-множественных отображений F_1: $\mathbf{R} \to P(\mathbf{R})$ и F_2: $\mathbf{R} \to P(\mathbf{R})$. Отображение F_1 в точке x^0 пн. сн., но не пн. св.; F_2 пн. св. на x^0, но не пн. сн.

---

*) $P(X)$ обозначает здесь множество всех подмножеств множества X.

**) В этом определении мы следуем Хьюарду [41]. Зангвилл [86] называет полунепрерывное сверху в нашем смысле точечно-множественное отображение замкнутым.

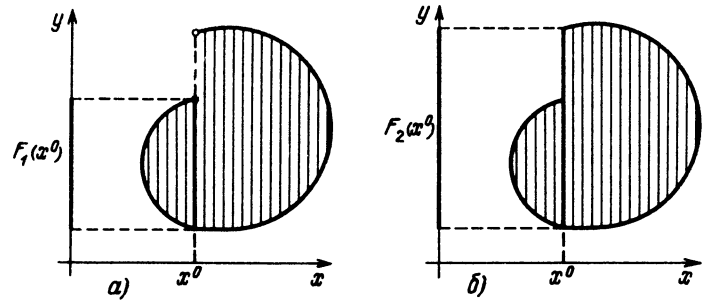


Рис. 6.25.

Пример 6.14. Любую функцию $f\colon X \to \mathbf{R}$ можно, очевидно, трактовать как точечно-множественное отображение. Если f как точечно-множественное отображение является пн.св. или пн. сн., то f – непрерывная функция.

Пример 6.15. Точечно-множественное отображение $F\colon X \times X \to P(X)$ вида $F(x, z) = [x, z]$, $x, z \in X$, непрерывно.

В самом деле, пусть $x^k \to x^0$, $z^k \to z^0$, $y^k \to y^0$ $(k \to \infty)$, где $y^k \in [x^k, z^k]$. Так как $y^k = \lambda_k x^k + (1 - \lambda_k) z^k$, $\lambda_k \in [0, 1]$, то $y^k - z^k = \lambda_k (x^k - z^k)$ и, тем самым, $\lim\limits_{k \to \infty} \lambda_k = \lambda_0$, $0 \leqslant \lambda_0 \leqslant 1$, т.е. $y^0 \in [x^0, z^0]$. Итак, отображение F пн. св.

Пусть теперь $x^k \to x^0$, $z^k \to z^0$, $y^0 \in [x^0, z^0]$. Тогда $y^0 = \lambda x^0 + (1 - \lambda) z^0$, $\lambda \in [0, 1]$. Если выбрать $y^k = \lambda x^k + (1 - \lambda) z^k$, то, очевидно, справедливо как $y^k \in F(x^k, z^k)$ $\forall k \in \mathbf{N}$, так и $y^k \to y^0$. Поэтому отображение F пн.сн. и, следовательно, непрерывно.

Точно так же показывается, что точечно-множественное отображение $F\colon X \times X \to P(X)$ вида

$$F(x, z) = \{y\colon\ y = x + \mu z,\quad \mu \geqslant 0\},\quad x, z \in X,\quad z \neq 0,$$

является непрерывным.

Мы ограничим наше дальнейшее изложение алгоритмами спуска, т.е. такими, для которых

$$f(x^{k+1}) \leqslant f(x^k) \quad \forall k \in \mathbf{N},\quad x^k \in X. \tag{6.155}$$

Пусть далее $G_1 = G$. Принимая во внимание условия $K1$ и $K2$ из § 6.1, естественно потребовать, чтобы

$$\begin{aligned} &x \notin G^0 \Rightarrow f(y) < f(x) \quad \forall y \in Q(x), \\ &x \in G^0 \Rightarrow f(y) \leqslant f(x) \quad \forall y \in Q(x). \end{aligned} \tag{6.156}$$

Теорема 6.29. *Пусть (Q, G) – алгоритм решения задачи P, обладающий следующими свойствами:*

а) *отображение Q пн. св. на $X \setminus G^0$;*

б) *выполняется условие (6.156), функция $f\colon \mathbf{R}^n \to \mathbf{R}$ непрерывна;*

в) *существует компактное множество G_2 такое, что $G \subseteq G_2 \subset X$ и $x^k \in G_2$ $\forall k \in \mathbf{N}$.*

Тогда любая предельная точка последовательности $\{x^k\}$ является решением задачи P.

Доказательство. В силу условия в) существует множество номеров K такое, что

$$\lim_{\substack{k\to\infty\\ k\in K}} x^k = x^* \in G_2, \quad \lim_{\substack{k\to\infty\\ k\in K}} x^{k+1} = x^{**} \in G_2.$$

Вследствие свойства а) имеем $x^{**} \in Q(x^*)$. Далее, из б) следует

$$f(x^k) \geqslant f(x^{k+1}) \geqslant f(x^l) \quad \forall l > k.$$

Отсюда с учетом непрерывности функции f получаем

$$f(x^*) \geqslant f(x^{**}) \geqslant f(x^*)$$

и, следовательно, $f(x^*) = f(x^{**})$.

Если допустить, что $x^* \notin G^0$, то в силу $x^{**} \in Q(x^*)$ из (6.156) имеем $f(x^{**}) < f(x^*)$. Последнее противоречит предыдущим выкладкам. Таким образом, $x^* \in G^0$.

Замечание. Вместо целевой функции f в соотношении (6.156) можно использовать произвольную непрерывную функцию, которая называется *функцией спуска*. Утверждение теоремы 6.29 справедливо и в этом случае [56, 86].

Чтобы применить теорему сходимости (теорему 6.29) к конкретным алгоритмам, целесообразно представить точечно-множественное отображение Q в виде композиции нескольких последовательных отображений. Определим отображение

$$F = F_2 \circ F_1, \quad F\colon\ X \to P(Z), \quad F_1\colon\ X \to P(Y), \quad F_2\colon\ Y \to P(Z), \tag{6.157}$$

где (рис. 6.26)

$$F(x) = \bigcup_{y \in F_1(x)} F_2(y). \tag{6.158}$$

Следующая теорема дает сведения о том, как свойства непрерывности отображений F_1 и F_2 переносятся на F.

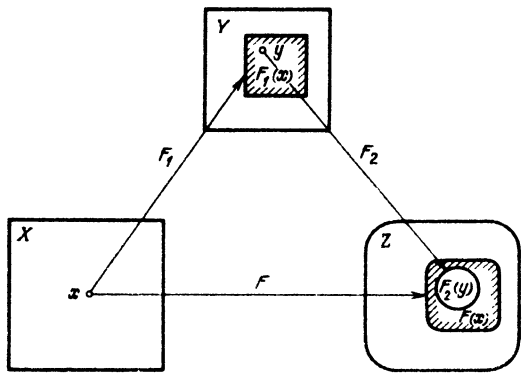


Рис. 6.26

Т е о р е м а 6.30. *Пусть* $F_1\colon X \to P(Y)$ *и* $F_2\colon Y \to P(Z)$. *Справедливы следующие утверждения*:

1) *если отображение* F_1 *пн.св. в точке* x^0, *отображение* F_2 *пн. св. на множестве* $F_1(x^0)$ *и выполняется соотношение*

$$\left.\begin{array}{l} y^k \to x^0 \\ y^k \in F_1(x^k) \end{array}\right\} \Rightarrow \left\{\begin{array}{l} \exists\, y^0 \in Y, \quad \exists K\colon \\ \lim\limits_{\substack{k\to\infty \\ k\in K}} y^k = y^0, \end{array}\right. \tag{6.159}$$

то отображение $F = F_2 \circ F_1$ *пн. св. в точке* x^0;

2) *если* F_1 *пн. сн. в точке* x^0, *а* F_2 *пн. сн. на множестве* $F_1(x^0)$, *то* $F = F_2 \circ F_1$ *пн. сн. в точке* x^0.

Д о к а з а т е л ь с т в о. Докажем лишь первое утверждение. Пусть $x^k \to x^0$, $z^k \to z^0$ $(k \to \infty)$, $z^k \in F(x^k)$. Требуется показать, что $z^0 \in F(x^0)$. Выберем $y^k \in F_1(x^k)$ таким образом, чтобы $z^k \in F_2(y^k)$. Тогда в силу (6.159) найдется множество номеров K такое, что $\lim\limits_{\substack{k\to\infty \\ k\in K}} y^k = y^0$. Тогда и $\lim\limits_{\substack{k\to\infty \\ k\in K}} z^k = z^0$. Так как F_1 и F_2 пн. св., то $y^0 \in F_1(x^0)$, $z^0 \in F_2(y^0)$ и, следовательно, $z^0 \in F_2 \circ F_1(x^0) = F(x^0)$.

З а м е ч а н и е. Справедливы импликации

$$Y \text{ компактно} \Rightarrow \text{выполняется (6.159)}, \tag{6.160}$$

$$\left.\begin{array}{l} F_1 - \text{непрерывное точечно-} \\ \text{множественное отображение} \end{array}\right\} \Rightarrow \text{выполняется (6.159)}. \tag{6.161}$$

Как известно из §§ 6.2 и 6.4, многие методы содержат алгоритмы решения специальных задач, таких, например, как минимизация функции на одномерном множестве или решение задачи линейного программирования. Поэтому представляют интерес свойства следующего точечно-множественного отображения:

$$B\colon U \times \hat{\mathbf{R}}_+ \to P(V), \quad U \subseteq \mathbf{R}^p, \quad V \subseteq \mathbf{R}^q, \quad \hat{\mathbf{R}}_+ \subseteq \mathbf{R}_+,$$

$$B(u, \epsilon) = \{v\colon v \in A(u), \quad \alpha(u, v) \leqslant \beta(u) + \epsilon\}, \tag{6.162}$$

где $\alpha\colon U \times V \to \mathbf{R}$, $\beta\colon U \to \mathbf{R}$,

$$\beta(u) = \inf_{v \in A(u)} \alpha(u, v), \quad A\colon U \to P(V). \tag{6.163}$$

Т е о р е м а 6.31. *Пусть* $A(u) \neq \phi$, $B(u, \epsilon) \neq \phi$ $\forall (u, \epsilon) \in U \times \hat{\mathbf{R}}_+$. *Справедливы следующие утверждения*:

1) *если множество* U *компактно, отображение* A *пн.св. в точке* u^0, *функция* α *пн.сн. на множестве* $A(u^0) \times \{u^0\}$, *то функция* β *пн.сн. в точке* u^0;

2) *если отображение* A *пн.св. в точке* u^0, *функция* α *пн.св. на множестве* $A(u^0) \times \{u^0\}$, *то функция* β *пн.св. в точке* u^0;

3) *если отображение* A *непрерывно в* u^0 *и функция* α *непрерывна на* $A(u^0) \times \{u^0\}$, *то отображение* $B(u, \epsilon)$ *пн.св. на множестве* $\{u^0\} \times \hat{\mathbf{R}}_+$.

Доказательство. 1. Пусть $\{u^k\} \subseteq U$ и $u^k \to u^0$ при $k \to \infty$. В силу (6.163) выполняется соотношение

$$\forall \delta > 0 \quad \exists v^k(\delta) \in A(u^k): \quad \beta(u^k) \geqslant \alpha(u^k, v^k(\delta)) - \delta. \tag{6.164}$$

Ввиду компактности множества U для любого $\delta > 0$ существует подпоследовательность $\{v^l(\delta)\}$ из $\{v^k(\delta)\}$ такая, что $\lim\limits_{l \to \infty} v^l(\delta) = v^0(\delta)$. Так как отображение A пн.св. в точке u^0, то $v^0(\delta) \in A(u^0)$ и с учетом (6.164) для всех $\delta > 0$ имеем

$$\varlimsup_{u^k \to u^0} \beta(u^k) \geqslant \varlimsup_{\substack{v^l(\delta) \to v^0(\delta) \\ u^k \to u^0}} \alpha(u^k, v^l(\delta)) - \epsilon \geqslant$$

$$\geqslant \alpha(u^0, v^0(\delta)) - \delta \geqslant \beta(u^0) - \delta. \tag{6.165}$$

Поскольку соотношения (6.165) выполняются для произвольных последовательностей $\{u^k\}$ таких, что $u^k \to u^0$ $(k \to \infty)$, то этим доказана полунепрерывность снизу функции β в точке u^0.

2. Пусть $\{u^k\} \subseteq U$ и $u^k \to u^0$ при $k \to \infty$. В силу (6.163) справедливо соотношение

$$\forall \delta > 0 \ \exists v^0(\delta): \quad \alpha(u^0, v^0(\delta)) \leqslant \beta(u^0) + \delta, \quad v^0(\delta) \in A(u^0). \tag{6.166}$$

Так как отображение A пн.сн., то для любого $\delta > 0$ существует последовательность $\{v^k(\delta)\}$ такая, что $v^k(\delta) \to v^0(\delta)$ при $k \to \infty$ и $v^k(\delta) \in A(u^k)$ $\forall k \geqslant k_0$. Поэтому

$$\beta(u^k) \leqslant \alpha(u^k, v^k(\delta)), \quad v^k(\delta) \in A(u^k) \quad \forall k \geqslant k_0,$$

и, с учетом (6.166)

$$\varlimsup_{u^k \to u^0} \beta(u^k) \leqslant \varlimsup_{\substack{u^k \to u^0 \\ v^k(\delta) \to v^0(\delta)}} \alpha(u^k, v^k(\delta)) \leqslant$$

$$\leqslant \alpha(u^0, v^0(\delta)) \leqslant \beta(u^0) + \delta \quad \forall \delta > 0. \tag{6.167}$$

Поскольку неравенства (6.167) выполняются для произвольных последовательностей $\{u^k\}$, сходящихся к u^0, отсюда следует полунепрерывность сверху функции β в точке u^0.

3. Если A непрерывно в u^0, а α непрерывна на $A(x^0) \times \{u^0\}$, то из первых двух утверждений теоремы вытекает, что функция β непрерывна в точке u^0 (при этом вследствие полунепрерывности снизу отображения A не требуется условия компактности U). Пусть $u^k \to u^0$, $\epsilon_k \to \epsilon_0$, $v^k \to v^0$ $(k \to \infty)$, $v^k \in B(u^k, \epsilon_k)$. Тогда $v^k \in A(u^k)$, $\alpha(u^k, v^k) \leqslant \beta(u^k) + \epsilon_k$ и, далее, $v^0 \in A(u^0)$, $\alpha(u^0, v^0) \leqslant \beta(u^0) + \epsilon_0$, а это означает, что $v^0 \in B(u^0, \epsilon_0)$.

Вслед за этими предварительными сведениями о точечно-множественных отображениях мы в качестве примера применения теоремы 6.29 докажем сходимость метода Коши (см. § 6.2) и одного метода возможных направлений (метода Франка – Вулфа).

В основе метода Коши лежат соотношения (6.5) и (6.6). Предположим, что $X = \mathbf{R}^n$, функция $f: \mathbf{R}^n \to \mathbf{R}$ выпукла и непрерывно дифференцируема,

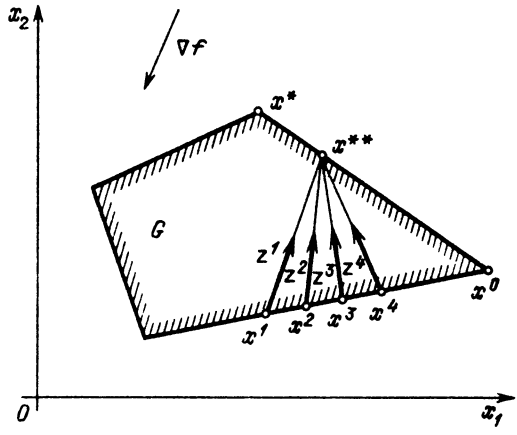


Рис. 6.27.

множество $G_2 = \{x \in \mathbf{R}^n\colon f(x^1) \geqslant f(x)\}$ компактно. Определим отображения $F\colon \mathbf{R}^n \to \mathbf{R}^{2n}$ и $B\colon \mathbf{R}^{2n} \to P(\mathbf{R}^n)$ следующим образом:

$$F(x) = (x, -\nabla f(x))',$$

$$B(x, z) = \{y\colon y = x + rz,\ r \geqslant 0;\ \min_{r \geqslant 0} f(x + rz) = f(y)\}.$$

В силу приведенных выше предположений F является непрерывным точечно-точечным отображением, а B при $z \neq 0$ будет, согласно теореме 6.31, полунепрерывным сверху точечно-множественным отображением (при этом следует учитывать, что в силу примера 6.15 отображение $A\colon \mathbf{R}^{2n} \to P(\mathbf{R}^n)$ вида $A(x, z) = \{y\colon y = x + rz, r \geqslant 0\}$ при $z \neq 0$ есть непрерывное точечно-множественное отображение и что $B(x, z) \neq \phi \quad \forall (x, z)' \in \in \mathbf{R}^{2n},\ z \neq 0$).

По теореме 6.30 и в силу (6.161) отображение $Q = B \circ F$ полунепрерывно сверху для $y \neq x$. Далее, для $x \notin G^0 = \{x\colon \nabla f(x) = 0\}$ справедливо неравенство

$$\min_{r \geqslant 0} f(x - r\nabla f(x)) < f(x),$$

т.е. имеет место условие (6.156). Следовательно, все предположения теоремы 6.29 выполнены, и этим доказана сходимость метода Коши.

Теперь покажем сходимость метода возможных направлений на примере алгоритма, приведенного в § 6.5 (стр. 228), для частного случая $I^L = I$. Далее, мы предположим, что число ϵ выбрано настолько большим, что для всех x^k выполняется $I_\epsilon^L(x^k) = I$. Пусть функция f выпукла и непрерывно дифференцируема, допустимая область $G = \{x\colon g_i(x) \leqslant 0,\ i \in I\}$ компактна и $C \supseteq G$. Тогда задача $P^2_{C,k}$, образованная согласно (6.78), примет вид

$$\min\{\nabla f(x^k)'z,\ z \in \tilde{G}\},$$

$$\tilde{G} = \{z\colon \nabla g_i(x^k)'z + g_i(x^k) \leqslant 0,\ i \subset I\}.$$

Алгоритм на стр. 228 совпадает в этом частном случае с *методом Франка – Вулфа* для решения задач выпуклого программирования с линейными ограничениями (см., например, работы [5, 37]).

На шаге 5 алгоритма требуется минимизировать функцию f на одномерном множестве. Если мы определим соответствующее отображение B точно так же, как в методе Коши, с дополнительным условием $y \in G$:

$$B(x, z) = \{ y:\ y = x + \lambda z,\ \lambda \geqslant 0;\ y \in G,\ \min_{\lambda \geqslant 0} f(x + \lambda z) = f(y) \},$$

то B: $G \times \mathbf{R}^n \to \boldsymbol{P}(G)$ не будет полунепрерывным сверху (см. в этой связи также работы [56, 86]). Данный факт поясняется на рис. 6.27. Пусть последовательности $\{x^k\}$ и $\{z^k\}$ таковы, что $x^k \to x^0$, $z^k \to (x^{**} - x^0)$ при $k \to \infty$. Очевидно,

$$y^k \in B(x^k, z^k) = \{ x^{**} \} \quad \forall k \geqslant 1,$$

но $\{ x^* \} = B(x^0, x^{**} - x^0)$ и $x^* \neq x^{**}$.

Для следующего отображения B: $G \times \tilde{G} \to \boldsymbol{P}(G)$ все же удается с помощью теоремы 6.31 показать полунепрерывность сверху:

$$B(x, u) = \{ y:\ y = x + \lambda u,\ \lambda \in [0, 1];\ \min_{\lambda \in [0, 1]} f(x + \lambda u) = f(y) \}.$$

Если мы, далее, определим отображение F: $G \to \boldsymbol{P}(G \times \tilde{G})$ как

$$F(x) = \{ x \} \times \{ u \in \tilde{G}:\ \min_{z \in \tilde{G}} \nabla f(x)'z = \nabla f(x)'u \},$$

то из теоремы 6.31 будет следовать, что F полунепрерывно сверху. Так как множество $G \times \tilde{G}$ компактно, а отображения F и B полунепрерывны сверху, то отображение $Q = B \circ F$ полунепрерывно сверху. Далее, из $x \notin G^0$ согласно теореме 6.10′ следует неравенство $\delta_k < 0$, а отсюда, в свою очередь, вытекает, что любой вектор $u \neq 0$, для которого $(x, u) \in F(x)$, будет подходящим направлением. Итак, $f(y) < f(x)$ $\forall y \in Q(x), y \neq x$, и тем самым справедливо условие (6.156). Следовательно, все предположения теоремы 6.29 выполнены, что влечет сходимость метода Франка – Вулфа.

1. *Abadie J.* On the Kuhn-Tucker Theorem. – In: Abadie J. (ed.) Nonlinear Programming. Amsterdam; North-Holland Publ. Comp., 1967, 19–36.
2. *Abdy P.R., Dempster M.A.H.* Introduction to Optimization Methods. – London: Chapman and Hall, 1974.
3. *Avriel M., Wilde D.J.* Optimality proof for the symmetric Fibonacci search technique. – Fibonacci Quart, 1966, 4, 265–269.
4. *Bachmann G., Elster K.H., Petry K.* Zur Problemstellung der geometrischen Optimierung. – Wiss. Z. TH Ilmenau, 1973, 19, 1, 3–39.
5. *Berge C., Ghouila-Houri A.* Programme, Spiele, Transportnetze. – Leipzig: Teubner, 1969.
6. *Bittner L.* Eine Verallgemeinerung des Verfahrens des logarithmischen Potentials von Frisch fur nichtlineare Optimierungsprobleme. – In: Prékopa A. (ed.). Colloq. on Applic. Math. to Econ. (Budapest 1963). Budapest: 1965. 43–53.
7*. *Céa J.* Optimisation: théorie et algorithmes. – Paris: Dunod, 1971.
8. *Collatz L., Wetterling W.* Optimierungsaufgaben. – Berlin–Heidelberg–New York: Springer-Verlag, 1966.
9. *Dantzig G.B., Eisenberg E., Cottle R.W.* Symmetric dual nonlinear programs. Pacific J. Math., 1965, 15, 809–812.
10. *Donath G.* Über direkte Lösungsverfahren der nichtlinearen Optimierung. – Diss. (A), TH Ilmenau 1975.
11. *Donath G., Elster K.-H.* Über Eigenschaften unimodaler Funktionen. – Wiss. Z. TH Ilmenau, 1972, 18, 3, 103–120.
12. *Donath G., Elster K.-H.* Über eine Verallgemeinerung unimodaler Funktionen. – Godišnik na višite techničeski učebni zavedenija (Matematika), 1973, IX, 3, 7–21.
13. *Dorn W.S.* A symmetric theorem for convex programs. – J. Operations Res. Soc. Japan, 1960, 2, 93–97.
14. *Dorn W.S.* Duality in quadratic programming. – Quart. Appl. Math., 1960, 18, 407–413.
15. *Elster K-H.* Ergebnisse und Probleme der nichtlinearen Optimierung. – Wiss. Z. TH Ilmenau, 1969, 15, 3, 37–61.
16. *Elster K-H., Folgmann G.* Über Verallgemeinerungen konvexer Funktionen und deren Anwendung in der Theorie der nichtlinearen Optimierung. – Wiss. Z. TH Ilmenau, 1970, 16, 4, 23–34.
17. *Elster K-H., Folgmann G.* Verallgemeinerte konvexe Funktionale in linearen Raumen. – Wiss. Z. TH Ilmenau Teil 1: 1974, 20, 3, 11–30; Teil 2: 1976, 22, 7, 55–83; Teil 3: 1977, 23, 4, 103–128.
18. *Elster K-H., Götz R.* Über die Constraint Qualification und damit verwandte Bedingungen. – Wiss. Z. TH Ilmenau, 1969, 15, 2, 27–35.
19. *Elster K-H., Grossmann C.* Untersuchungen zu Regularitätsbedingungen bei den Barriere- und Zentrenmethoden. – Math. Operationsforschung und Statistik, 1974, 5, 3, 191–206.
20. *Elster K-H., Grossmann C., Rathgen W.* Lineare Approximationsverfahren in der nichtlinearen Optimierung (Teil 2). – Wiss. Z. TH Ilmenau, 1972, 18, 2, 134–155.
21. *Elster K-H., Kempe W.* Neuere Ergebnisse in der nichtlinearen Optimierung. – Wiss. Z. TH Ilmenau, 1974, 20, 4/5, 33–59.
22. *Elster K-H., Nehse R.* Zum Dualitätssatz von Fenchel. – Math. Operationsforschung und Statistik, 1974, 5, 4/5, 269–280.
23. *Elster K-H., Suppe C.* Zur Entwicklung der Dualitätstheorie in der nichtlinearen Optimierung. – Wiss. Z. TH Ilmenau, 1969, 15, 4/5, 77–98.

24. *Erdmann D.* Zur Entwicklung der Theorie der Minimaxsätze. – Diss. (A), Ilmenau, 1974.
25. *Fenchel W.* On conjugate convex functions. – Canad. J. Math. 1949, 1, 73–77.
26. *Fenchel W.* Convex Cones, Sets and Functions. Princeton: Princeton Univ. Press, 1953.
27* *Fiacco A.V., McCormick G.P.* Nonlinear Programming. Sequential Unconstrained Minimization Techniques. – New York: Wiley, 1968.
28* *Fichtenholz G.M.* Differential – und Integralrechnung (Band I) Berlin: VEB Deutscher Verlag der Wissenschaften, 1964.
29. *Fletcher R.* (ed.) Optimization New York: Academic Press, 1969.
30. *Focke J.* Eine Modifikation des reduzierten Gradientenverfahrens. – Math. Operationsforschung und Statistik, 1975, 6, 4, 561–570.
31. *Folgmann G.* Über Verallgemeinerung knovexer Funktionen in linearen Raumen. – Diss. (A), PH Halle, 1974.
32. *Geoffrion A.M.* Duality in nonlinear programming: A simplified applicationsoriented development. – SIAM Review, 1971, 13, 1, 1–37.
33* *Golstein E.G.* Konvexe Optimierung. WTB, Bd. 122. – Berlin: Akademie-Verlag, 1973.
34* *Golstein E.G.* Dualitätstheorie in der nichtlinearen Optimierung und ihre Anwendung. – Berlin: Akademie-Verlag, 1975.
35. *Gopfert A.* Mathematische Optimierung in allgemeinen Vektorräumen. – Leipzig: Teubner, 1973.
36. *Grossmann C.* Straf- und Barrieremethoden in der nichtlinearen Optimierung. – Diss. (A), TH Ilmenau, 1972.
37* *Hadley G.* Nichtlineare und dynamische Programmierung. – Berlin: Verlag Die Wirthshaft, 1969.
38. *Hanson M.A.* A duality theorem in nonlinear programming with nonlinear constraints. – Austr. J. of Statist., 1961, 3, 64–72.
39. *Hestenes M.R.* Elements of the Calculus of Variations. – New York: McGraw-Hill, 1956.
40* *Himmelblau D.M.* Applied Nonlinear Programming. – New York: McGraw-Hill, 1972.
41. *Huard P.* Optimization algorithms and point-to-set-maps. – Math. Progr., 1975, 8, 3, 308–331.
42. *John F.* Extremum problems with inequalities as subsidary conditions. – In: Studies and essays. Courant anniversary volume. New York: Interscience Publ. Inc., 1948, 187–204.
43. *Karamardian S.* Duality in mathematical programming. – J. Math. Anal. and Appl., 1967, 20, 344–358.
44. *Kelley J.E.* Jr. The cutting plan method for solving convex programs. – J. Soc. Indust. Appl. Math., 1960, 9, 703–712.
45. *Kleibohm K.* Un procédé pour la resolution approximative d'un programme convexe. – C.R. Acad. Sci. Paris, 1965, 261, 306–307.
46. *Kleinmichel H.* Über die Gewinnung von Optimalitätsbedingungeñ für nichtkonvexe Optimierungsaufgaben durch Lösung linearer Programme. – ZAMM, 1971, 51, 587–594.
47. *Kleinmichel H.* Zur Konvergenz der Verfahren der zulässigen Richtungen. – Beitrage zur Numerischen Mathematik, 1975, 4, 115–127.
48* *Korbut A.A., Finkelstein I.I.* Diskrete Optimierung.– Berlin: Akademie-Verlag, 1971.
49. *Krekó B.* Optimierung, Nichtlineare Modelle.– Berlin: VEB Deutscher Verlag der Wissenschaften, 1974.
50. *Kuhn H.W., Tucker A.W.* Nonlinear programming. – Proc. of the Second Berkeley Symp. on Math. Stat. and Prob., Berkeley, Calif., 1951, 481–492.
51. *Künzi H.P., Oettli W.* Nichtlineare Optimerung, Neuere Verfahren, Bibliographie (Lecture Notes Op. Res. 16). – Berlin – Heidelberg – New York: Springer-Verlag, 1969.
52. *Künzi H.P., Tzschach H.G., Zehnder C.A.* Numerische Methoden der Mathematischen Optimerung mit ALGOL- und FORTRAN-Programmen. – Leipzig: Teubner, 1969.
53* *Lasdon L.S.* Optimization Theory for Large Systems. – London: Macmillan, 1971.
54. *Lootsma T.A.* (ed.) Numerical Methods for Non-linear Optimization. – New York: Academic Press, 1972.
55. *Luenberger D.G.* Optimization by Vector Space Methods. – New York: Wiley, 1969.
56. *Luenberger D.G.* Introduction to Linear and Nonlinear Programming. – Reading Mass., Addison -Wesley, 1973.

57. *Mangasarian O.L.* Pseudo-convex functions. – J. SIAM Control., Ser. A, 1965, 3, 2, 281–290.
58. *Mangasarian O.L., Ponstein J.* Minimax and duality in nonlinear programming. – J. Math. Anal. Appl., 1965, 11, 504–518.
59. *Mangasarian O.L.* Nonlinear Programming. – New York: McGraw-Hill, 1969.
60. *Mangasarian O.L.* Optimality and duality in nonlinear programming. – In: Proc. of the Princeton Sympos. on Math. Program. ed. by H.W. Kuhn, Princeton Univ. Press, 1970, 429–443.
61. *Martos B.* Nonlinear Programming. Theory and Methods. – Budapest: Akadémie Kiadó, 1975.
62. *Moreau J.J.* Convexity and duality. – In: Caianiello E.R. (ed.) Functional Analysis and Optimization. New York: Academic Press, 1966, 145–169.
63. *Neumann J.* von. Zur Theorie der Gesellschaftsspiele. – Math. Ann., 1928, 100, 295–320.
64. *Nožička F., Guddat J., Hollatz H., Bank B.* Theorie der linearen parametrischen Optimierung. – Berlin: Akademie-Verlag, 1974.
65. *Oliver L.T., Wilde D.J.* Symmetric sequential minimax search for a maximum. – Fibonacci Quart., 1966, 4, 163–175.
66* *Ortega J.M., Rheinboldt W.C.* Iterative Solution of Nonlinear Equations in Several Variables. – New York: Academic Press, 1970.
67. *Petry K.* Beiträge zur geometrischen Optimierung. – Diss. (A), TH Ilmenau, 1974.
68. *Pironneau O., Polak E.* Rate of convergence of a clas of methods of of feasible directions. – SIAM J. Numer. Anal., 1973, 10, 1, 161–174.
69* *Polak E.* Computational Methods in Optimization. A Unified Approach.– New York: Academie Press, 1971.
70. *Ponstein J.* Seven kinds of convexity. – SIAM Rev., 1967, 9, 115–119.
71* *Pšeničnyj B.N.* Notwendige Optimalitätsbedingungen. – Leipzig: Teubner, 1972.
72. *Pun L.* Abriss der Optimierungspraxis. – Berlin: Akademie-Verlag, 1974.
73. *Reinhardt R.* Dualität und Sattelpunktprobleme in der nichtlinearen Optimierung. – Diss. (A) TH Ilmenau, 1971.
74* *Rockafellar R.* Convex Analysis. Princeton: Princeton Univ. Press, 1970.
75. *Sadowski H.* Untersuchungen zu den Verfahren der reduzierten Gradienten in der nichtlinearen Optimierung. – Diss. (A), TU Dresden, 1973.
76. *Seidler K.H.* Zur Dualisierung in der nichtlinearen Optimierung.– Diss. (A), TH Ilmenau, 1972.
77. *Stoer J.* Duality in nonlinear programming and the minimax-theorem. – Num. Math., 1963, 5, 371–379.
78. *Stoer J.* Über einen Dualitätssatz in der nichtlinearen Programmierung. – Num. Math., 1964, 6, 55–58.
79. *Stoer J., Witzgall Ch.* Convexity and Optimization in Finite Dimensions I. – Berlin – Heidelberg – New York: Springer-Verlag, 1970.
80. *Topkis D.M.* Cutting-plane methods without nested constraint sets.– Operations Res., 1970, 18, 404–413.
81* *Tschernikow S.N.* Lineare Ungleichungen.– Berlin: VEB Deutscher Verlag der Wissenschaften, 1971.
82. *Valentine F.A.* Convex Sets. – New York: McGraw-Hill Book Comp., 1964.
83. *Vogel W.* Duale Optimierungsaufgaben und Sattelpunktsätze. – Unternehmensforschung, 1969, 13, 1–28.
84. *Vogel W.* Lineares Optimieren.– Leipzig: Akademische Verlagsgesellschaft Geest & Portig K.G., 1970.
85* *Wolfe P.* A duality theorem for nonlinear programming. – Quart. Appl. Math., 1961, 19, 239–244.
86. *Zangwill W.I.* Nonlinear programming. A Unified Approach. – Englewood Cliffs (N.J.), Prentice-Hall, 1970.
87* *Zoutendijk G.* Methods of Feasible Directions. A Study in Linear and Non-linear Programming. – Amsterdam: Elsevier, 1960.
88. *Беленький В.З., Волконский В.А., Иванков С.А., Поманский А.Б., Шапиро А.Д.* Итеративные методы теории игр и программирования. – М.: Наука, 1974.
89. *Васильев Ф.П.* Лекции по методам решения экстремальных задач. – Изд. МГУ, 1974.

90. *Гирсанов И.В.* Лекции по математической теории экстремальных задач. – Изд. МГУ, 1970.
91. *Гольштейн Е.Г., Третьяков Н.В.* Модифицированные функции Лагранжа. – Экономика и матем. методы., 1974, X, 3, 568–591.
92. *Демьянов В.Ф., Малоземов В.Н.* Введение в минимакс. – М.: Наука, 1972.
93. *Демьянов В.Ф., Рубинов А.М.* Приближенные методы решения экстремальных задач. – Изд. ЛГУ, 1968.
94. *Дубовицкий А.Я., Милютин А.А.* Задача на экстремум при наличии ограничений. – ЖВМ, 1965, 5, 3, 395–453.
95. *Зуховицкий С.И., Авдеева Л.И.* Линейное и выпуклое программирование. – М.: Наука, 1965.
96. *Иоффе А.Д., Тихомиров В.М.* Теория экстремальных задач. – М.: Наука, 1974.
97. *Карманов В.Г.* Математическое программирование. – М.: Наука, 1975.
98. *Левитин Е.С., Поляк Б.Т.* Методы минимизации при наличии ограничений. – ЖВМ, 1966, 6, 5, 787–823.
99. *Любич Ю.И., Майстровский Г.Д.* Общая теория релаксационных процессов для выпуклых функционалов. – УМН, 1970, 25, 57–112.
100. *Поляк Б.Т.* Градиентные методы минимизации функционалов. – ЖВМ, 1963, 3, 4, 643–654.
101. *Пшеничный Б.Н., Данилин Ю.Н.* Численные методы в экстремальных задачах. – М.: Наука, 1975.

Список цитируемых работ, опубликованных на русском языке*)

1. *Гольштейн Е.Г.* Выпуклое программирование. Элементы теории. – М.: Наука, 1970.
2. *Гольштейн Е.Г.* Теория двойственности в математическом программировании и ее приложения. – М.: Наука, 1971.
3. *Зангвилл У.И.* Нелинейное программирование. Единый подход. – М.: Сов. радио, 1973.
4. *Зойтендейк Г.* Методы возможных направлений. – М.: ИЛ, 1963.
5. *Корбут А.А., Финкельштейн Ю.Ю.* Дискретное программирование. – М.: Наука, 1969.
6. *Лэсдон Л.С.* Оптимизация больших систем. – М.: Наука, 1975.
7. *Ортега Дж., Рейнболдт В.* Итерационные методы решения нелинейных систем уравнений со многими неизвестными. – М.: Мир, 1975.
8. *Полак Э.* Численные методы оптимизации. Единый подход. – М.: Мир, 1974.
9. *Пшеничный Б.Н.* Необходимые условия экстремума. – М.: Наука, 1969.
10. *Рокафеллар Р.Т.* Выпуклый анализ. – М.: Мир, 1973.
11. *Сеа Ж.* Оптимизация. Теория и алгоритмы. – М.: Мир, 1973.
12. *Фиакко А., Мак-Кормик Г.* Нелинейное программирование. Методы последовательной безусловной минимизации. – М.: Мир, 1972.
13. *Фихтенгольц Г.М.* Курс дифференциального и интегрального исчисления. – М.: Наука, 1958, т. 1.
14. *Химмельблау Д.* Прикладное нелинейное программирование. – М.: Мир, 1975.
15. *Хедли Дж.* Нелинейное и динамическое программирование. – М.: Мир, 1967.
16. *Черников С.Н.* Линейные неравенства. – М.: Наука, 1968.

Дополнительный список работ по оптимизации, опубликованных на русском языке

1. *Аоки М.* Введение в методы оптимизации. – М.: Наука, 1977.
2. *Астафьев Н.Н.* Линейные неравенства и выпуклость. – М.: Наука, 1982.
3. *Базара М., Шетти К.* Нелинейное программирование. Теория и алгоритмы. – М.: Мир, 1982.

*) Названия этих работ в основном списке литературы помечены знаком *.

4. *Булавский В.А., Звягина Р.А., Яковлева М.А.* Численные методы линейного программирования. – М.: Наука, 1977.
5. *Булатов В.П.* Методы погружения в задачах оптимизации. – Новосибирск: Наука, 1977.
6. *Васильев Ф.П.* Численные методы решения экстремальных задач. – М.: Наука, 1980.
7. *Васильев Ф.П.* Методы решения экстремальных задач. – М.: Наука, 1981.
8. *Габасов Р., Кириллова Ф.М.* Методы оптимизации. – Минск: БГУ, 1975.
9. *Гермейер Ю.Б.* Введение в теорию исследования операций. – М.: Наука, 1971.
10. *Данциг Дж.* Линейное программирование, его применения и обобщения. – М.: Прогресс, 1966.
11. *Даффин Р., Петерсон Э., Зенер К.* Геометрическое программирование. – М.: Мир, 1972.
12. *Евтушенко Ю.Г.* Методы решения экстремальных задач и их применение в системах оптимизации. М.: Наука, 1982.
13. *Емеличев В.А., Ковалев М.М., Кравцов М.К.* Многогранники, графы, оптимизация. – М.: Наука, 1981.
14. *Еремин И.И., Астафьев Н.Н.* Введение в теорию линейного и выпуклого программирования. – М.: Наука, 1976.
15. *Еремин И.И., Мазуров В.Д.* Нестационарные процессы математического программирования. – М.: Наука, 1979.
16. *Еремин И.И., Мазуров В.Д., Астафьев Н.Н.* Несобственные задачи линейного и выпуклого программирования. – М.: Наука, 1983.
17. *Ермольев Ю.М.* Методы стохастического программирования. – М.: Наука, 1976.
18. *Иванов В.К., Васин В.В., Танана В.П.* Теория линейных некорректных задач и ее приложения. – М.: Наука, 1978.
19. *Интрилигатор М.Д.* Математические методы оптимизации и экономическая теория. – М.: Прогресс, 1975.
20. *Карлин С.* Математические методы в теории игр, программировании и экономике. – М.: Мир, 1964.
21. *Красовский Н.Н., Субботин А.И.* Позиционные дифференциальные игры. – М.: Наука, 1974.
22. *Кюнци Г.П., Крелле В.* Нелинейное программирование. – М.: Сов. радио, 1965.
23. Линейное и нелинейное программирование / *Ляшенко И.Н., Карагодова Е.А., Черникова Н.В., Шор Н.З.*; Под ред. И.Н.Ляшенко. – Киев: Вища школа, 1975.
24. Линейные неравенства и смежные вопросы / Под ред. Г. Куна, А. Таккера. – М.: ИЛ, 1959.
25. *Лоран П.-Ж.* Аппроксимация и оптимизация. – М.: Мир, 1975.
26. *Михалевич В.С., Кукса А.И.* Методы последовательной оптимизации в дискретных сетевых задачах оптимального распределения ресурсов. – М.: Наука, 1983.
27. *Моисеев Н.Н., Иванилов Ю.П., Столярова Е.М.* Методы оптимизации. – М.: Наука, 1978.
28. *Мухачева Э.А., Рубинштейн Г.Ш.* Математическое программирование. – Новосибирск: Наука, 1977.
29. *Немировский А.С., Юдин Д.Б.* Сложность задач и эффективность методов оптимизации. – М.: Наука, 1980.
30. *Никайдо Х.* Выпуклые структуры и математическая экономика. – М.: Мир, 1972.
31. *Нурминский Е.А.* Численные методы решения детерминированных и стохастических минимаксных задач. – Киев: Наукова думка, 1979.
32. *Поляк Б.Т.* Введение в оптимизацию. – М.: Наука, 1983.
33. *Понтрягин Л.С., Болтянский В.Г., Гамкрелидзе Р.В., Мищенко Е.Ф.* Математическая теория оптимальных процессов. – М.: Физматгиз, 1961.
34. *Пропой А.И.* Элементы теории оптимальных дискретных процессов. – М.: Наука, 1973.
35. *Пшеничный Б.Н.* Выпуклый анализ и экстремальные задачи. – М.: Наука, 1980.
36. *Романовский И.В.* Алгоритмы решения экстремальных задач. – М.: Наука, 1977.

37. *Стронгин Р.Г.* Численные методы в многоэкстремальных задачах. – М.: Наука, 1978.
38. *Тихонов А.Н., Арсенин В.Я.* Методы решения некорректных задач. – М.: Наука, 1974.
39. *Федоренко Р.П.* Приближенное решение задач оптимального управления. – М.: Наука, 1978.
40. *Федоров В.В.* Численные методы максимина. – М.: Наука, 1979.
41. Численные методы условной оптимизации / Под ред. Ф. Гилла, У. Мюррея. – М.: Мир, 1977.
42. *Шор Н.З.* Методы минимизации недифференцируемых функций и их приложения. – Киев: Наукова думка, 1979.
43. *Экланд И., Темам Р.* Выпуклый анализ и вариационные проблемы. – М.: Мир, 1979.
44. *Эрроу К.Дж., Гурвиц Л., Удзава Х.* Исследования по линейному и нелинейному программированию. – М.: ИЛ, 1962.

Алгоритм 185

Базис канонический 37

Векторы аффинно зависимые 24
– – независимые 24
– линейно зависимые 24
– – независимые 24
Внутренность 41
– относительная 44

Гиперплоскость вертикальная 71
– невертикальная 71
– несобственная опорная 57
– образующая 23
– опорная 56
– разделяющая 48
– собственная опорная 57
– – разделяющая 48
– строго разделяющая 48
Граница относительная 44

Длина шага 191

Задача без ограничений 16, 189
– выпуклого квадратичного программирования 16
– – программирования 16
– – – двойственная 168
– – – несовместная 168
– – – прямая 168
– – – совместная 168
– – – устойчивая 170
– гиперболического программирования 16
– двойственная 145
– квадратичного программирования 16
– линейного программирования 15
– максимизации 15
– математического программирования общая 13
– минимизации 15
– нелинейного программирования 15
– несовместная 14
– о седловой точке 140
– однородного программирования 17
– позиномиального программирования 17
– прямая 149
– с ограничениями в форме равенств 16
– сепарабельного программирования 17
– сильно совместная 172
Задача совместная 14
– устойчивая 153
– целочисленного программирования 15
Замыкание снизу 82

Комбинация линейная 24
– – аффинная 24
– – выпуклая 24
– – неотрицательная 24
– – положительная 24
Конус 29
– выпуклый 22
– касательный 121
– линеаризующий 124
– многогранный 23
– полярный 52
– противоположный 29

Луч 29

Метод барьеров 216
– внешнего штрафа 235
– внутреннего штрафа 235
– Коши 191
– Ньютона 193, 194
– приведенных градиентов 215
– проектируемых градиентов 215
– решения 186
– спуска 187
– Франка – Вульфа 255
– штрафов 215
Методы возможных направлений 214
– квазиньютоновские 194, 195
– секущих плоскостей 217
– сопряженных направлений 195, 196
Многогранник 33
Многообразие линейное 22
Множество выпуклое 22
– дважды сопряженное 101
– допустимое 14
– многогранное 23
– относительно открытое 44
– решений 14
– сопряженное 99
– – сверху 99
– – снизу 99
– трижды сопряженное 101
Множитель Лагранжа 20
– сходимости 188

Надграфик 61
Направление движения 191

Область допустимая 14
– эффективная 84
Оболочка аффинная 30
– выпуклая 30
– замкнутая 41
– коническая 30
– – выпуклая 30
– линейная 30
– положительная 30
Ограничение 15
– активное 123
Операция взятия оболочки 30
Отображение полунепрерывное сверху 249
– – снизу 249

Пара последовательностей золотого сечения 209
– – Фибоначчи 205
Подграфик 61
Подпространство 22
Полупространство опорное 57
Полупрямая 29
Последовательность, сходящаяся в данном направлении (направленно сходящаяся) 120
– – квадратично 188
– – линейно 188
– – сверхлинейно 188
Преобразование аффинное 26
– линейное 26

Размерность линейного многообразия 33
– множества 33
Разрыв в двойственности 146
Решение глобальное 14
– оптимальное 14

Свойство "верхней оценки" (*HVP*) 178
– "нижней оценки" (*LVP*) 178
Сдвиг 26
Симплекс m-мерный 33
Скалярное кратное выпуклого множества 26
Субградиент 72, 85
Субдифференциал 72, 85
Сумма выпуклых множеств 26
Сходимость алгоритма 186

Теорема Вейля 53
– двойственности сильная 151
– – слабая 145
– Джона 126
– Жордана об альтернативе 97
– Каратеодори 34
– – для концов 33
– Крейна – Мильмана 56
– Куна – Таккера 127
– Минковского 37
Теорема Моцкина об альтернативе 97
– о минимаксе 142
– о резрешимости 19
– о симметрии 104
– отделимости вторая 50
– – первая 49
– Фаркаша 53
– – для неоднородных систем 98
– фон Неймана о минимаксе 175
– Хелли 39
Точка локального минимума 14
– опорная 56
– относительно внутренняя 44
– – граничная 44
– седловая 139
– собственного локального минимума 14
– экстремальная 54
Точки сопряженные 109

Условие неотрицательности 16
– Хьюарда 167
– Хэнсона 167
Условия Куна – Таккера 127
– регулярности 115, 126, 127, 134, 137, 138

Функция аффинная 15
– вогнутая 58, 84
– выпуклая 58, 84
– – в общем смысле в точке 86
– – – – – на множестве 86
– замкнутая сверху 79
– – снизу 79
– квазивыпуклая 87
– квазиунимодальная 197
– Лагранжа 20, 127, 136
– ограничений 15
– положительно однородная степени 1 111
– полунепрерывная сверху в точке 65
– – – на множестве 65
– – снизу в точке 65, 84
– – – на множестве 65, 84
– почти выпуклая 93
– псевдовыпуклая 87
– собственная выпуклая 84
– сопряженная сверху 105
– – снизу 105
– спуска 251
– строго вогнутая 58
– – выпуклая 58
– – – в общем смысле 87
– – квазивыпуклая 87
– – – псевдовыпуклая 87
– – явно квазивыпуклая 93
– унимодальная 197
– чувствительности 146, 170
– явно квазивыпуклая 93

ВВЕДЕНИЕ
В НЕЛИНЕЙНОЕ
ПРОГРАММИРОВАНИЕ

Серия: "Экономико-математическая библиотека"

Редакторы А.Д. Вайнштейн, Е.Ю. Ходан
Технический редактор В.В. Лебедева
Корректоры Т.В. Обод, Е.А. Янышева

Набор осуществлен в издательстве
на наборно-печатающих автоматах

ИБ № 12377

Сдано в набор 04.01.85. Подписано к печати 05.04.85
Формат 60 X 90 1/16. Бумага офсетная № 1
Гарнитура Пресс-Роман. Печать офсетная
Усл. печ.л. 16,5. Усл.кр.-отт. 16,5. Уч.-изд.л. 17,21
Тираж 7500 экз. Тип. зак. 383 . Цена 2 руб.

Ордена Трудового Красного Знамени
издательство "Наука"
Главная редакция физико-математической литературы
117071 Москва, В-71, Ленинский проспект, 15

4-я типография издательства "Наука"
630077 г. Новосибирск-77, ул. Станиславского, 25